GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: December 24, 2002, 08:48:45 ; Search time 40 Seconds
(without alignments)
2368.530 Million cell updates/sec

Title: US-09-966-561-2
Perfect score: 3792
Sequence: 1 MAERESGGLGGGAASPPAAS.....QQFYKQFVEYTCPTEDIYLE 711

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3780	99.7	711	19	AAW80602	Human transcriptio
2	3544.5	93.5	708	22	AAG78707	Rat SAPK interacti
3	3544.5	93.5	715	22	AAG78708	Rat SAPK interacti
4	3529	93.1	707	22	AAG78825	Murine SKIP-1b. M
5	3503	92.4	714	19	AAW80601	Rat transcription
6	3389	89.4	699	22	AAG78706	Rat SAPK interacti
7	3367.5	88.8	698	22	AAG78826	Murine SKIP-2a. M
8	3262.5	86.0	660	20	AAW81525	Murine JNK-interac
9	3248	85.7	673	22	AAG78824	Murine SKIP-2b. M
10	3245	85.6	659	23	ABB04788	LDL receptor bindi
11	3104	81.9	617	22	AAG78827	Murine SKIP-3. Mu
12	2875	75.8	659	23	ABB04797	LDL receptor bindi
13	2864	75.5	659	23	ABB04793	LDL receptor bindi
14	2846	75.1	659	23	ABB04790	LDL receptor bindi
15	2845	75.0	659	23	ABB04794	LDL receptor bindi
16	2829	74.6	659	23	ABB04789	LDL receptor bindi
17	2824	74.5	659	23	ABB04791	LDL receptor bindi
18	2817	74.3	659	23	ABB04792	LDL receptor bindi
19	2816	74.3	659	23	ABB04795	LDL receptor bindi
20	2814	74.2	659	23	ABB04796	LDL receptor bindi
21	1428	37.7	299	20	AAW88550	Secreted protein e
22	1428	37.7	299	22	ABB50317	Human secreted pro
23	1279.5	33.7	830	23	ABB04808	LDL receptor bindi
24	1116.5	29.4	830	23	ABB04813	LDL receptor bindi
25	1111.5	29.3	830	23	ABB04810	LDL receptor bindi
26	1103.5	29.1	829	23	ABB04817	LDL receptor bindi
27	1097.5	28.9	830	23	ABB04815	LDL receptor bindi
28	1093.5	28.8	830	23	ABB04811	LDL receptor bindi
29	1083.5	28.6	830	23	ABB04809	LDL receptor bindi
30	1075.5	28.4	830	23	ABB04814	LDL receptor bindi
31	1074	28.3	830	23	ABB04816	LDL receptor bindi
32	1060.5	28.0	830	23	ABB04812	LDL receptor bindi
33	1056.5	27.9	598	20	AAW99065	Human U62317 prote
34	751	19.8	155	20	AAW81527	JIP-1 JNK binding
35	537.5	14.2	483	22	ABB57918	Drosophila melanog
36	392	10.3	79	20	AAW81537	JIP-1 polypeptide
37	389	10.3	77	20	AAW81538	JIP-1 polypeptide
38	359	9.5	76	20	AAW81536	JIP-1 polypeptide
39	327	8.6	68	20	AAW81539	JIP-1 polypeptide
40	305	8.0	59	20	AAW81540	JIP-1 polypeptide
41	261	6.9	49	20	AAW81541	JIP-1 polypeptide
42	260	6.9	50	20	AAW81535	JIP-1 SH3 domain.
43	254	6.7	66	22	ABB50572	Human secreted pro
44	207	5.5	39	20	AAW81542	JIP-1 polypeptide
45	206	5.4	1239	22	ABG09877	Novel human diagno

ALIGNMENTS

RESULT 1
AAW80602
ID AAW80602 standard; Protein; 711 AA.
XX
AC AAW80602;
XX
DT 15-FEB-1999 (first entry)
XX
DE Human transcription factor islet-brain 1 (IB1).
XX
KW IB1; islet-brain 1; transcription factor; human; diabetes;
KW dementia; Parkinson's disease; Alzheimer's disease; epilepsy;
KW neuroblastoma; glioblastoma; apoptosis; cancer; autoimmune disease;
KW systemic lupus erythematosus; myocardial infarction; ischaemia;
KW diagnosis; therapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 282
FT /note= "encoded by TAC"
XX
PN WO9844106-A1.
XX
PD 08-OCT-1998.
XX
PF 02-APR-1998; 98WO-GB00972.
XX
PR 15-MAY-1997; 97GB-0009920.
PR 03-APR-1997; 97GB-0006731.
XX
PA (KIDD/) KIDDLE S J.
PA (NICO/) NICOD P.

PA (WAEB/) WAEBER G.
XX
PI Bonny C, Waeber G;
XX
DR WPI; 1998-568278/48.
DR N-PSDB; AAV62463.
XX
PT New isolated transcription factor islet-brain 1 - used to develop
PT products for treating e.g. diabetes, neurodegenerative disorders,
PT cancers, autoimmune disease, heart disease or epilepsy
XX
PS Claim 2; Fig 1F; 111pp; English.
XX
CC This is the amino acid sequence of human islet-brain 1 (IB1), a
CC novel transcriptional activator that is involved in the control of
CC the GLUT2 and insulin genes by interacting with homologous
CC cis-regulatory elements of the GLUT2 and insulin gene promoters.
CC The amino acid sequence was deduced from the sequence of isolated
CC IB1 cDNA (see AAV80602) derived from human insulinoma INS-1 cells.
CC Rat IB1 polynucleotide (see AAV62462) and polypeptide (see AAW80601)
CC are also claimed. IB1 polypeptides, nucleic acids, agonists and
CC antagonists can be used in the treatment or diagnosis of diabetes,
CC neurodegenerative disorders or cancers, to inhibit cells from
CC undergoing apoptosis, to treat dementia, Parkinsonism, Alzheimer's
CC disease, neuronal disabilities such as speech disorders and memory
CC alteration, autoimmune diseases affecting the CNS such as systemic
CC lupus erythematosus, diabetes, heart diseases such as myocardial
CC infarct and ischemia or brain attack, neurological disorders, e.g.
CC neuroblastoma, glioblastoma or cancers, or to promote apoptosis in
CC cells, or treat refractory epilepsy. They can also be used for
CC screening test substances for IB1 biological activity. IB1
CC producing cells can be used therapeutically to produce IB1 in a
CC subject.
XX
SQ Sequence 711 AA;

Query Match 99.7%; Score 3780; DB 19; Length 711;
Best Local Similarity 99.7%; Pred. No. 4.4e-287;
Matches 709; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60

Qy     61 QCKDTLSLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAAR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 QCKDTLSLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAAR 120

Qy    121 RPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNNN 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 RPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNNN 180

Qy    181 SLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRRS 240
          ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db    181 SLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSDELPPQSGPAPTTDRGTSTDSPCRRS 240

Qy    241 TATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAFL 300
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    241 TATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEINLTPVQRPPDAAEPTSAFL 300

Qy    301 PPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEPP 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 PPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEPP 360

Qy    361 PPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASVS 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 PPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASVS 420

Qy    421 SPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLF 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 SPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLF 480

Qy    481 SCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFP 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFP 540

Qy    541 AYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRLT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRLT 600

Qy    601 VHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGFI 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 VHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGFI 660

Qy    661 TKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
```

RESULT 2
AAG78707
ID AAG78707 standard; Protein; 708 AA.
XX
AC AAG78707;
XX
DT 11-DEC-2001 (first entry)
XX
DE Rat SAPK interacting protein #2.
XX
KW Rat; SKIP; SAPK interacting protein; stress-activated kinase;
KW cancer; inflammation; apoplexy; cell death; apoptosis.
XX
OS Rattus norvegicus.
XX

FH	Key	Location/Qualifiers
FT	Domain	126..283
FT		/label= JNK_binding_domain
FT	Domain	366..370
FT		/label= proline_rich_domain
FT	Domain	484..546
FT		/label= SH3_domain
FT	Domain	529..594
FT		/label= helix-loop-helix_motif
FT	Domain	563..697
FT		/label= phosphotyrosine_interaction_domain

XX
PN KR2001029353-A.
XX
PD 06-APR-2001.
XX
PF 30-SEP-1999; 99KR-0042119.
XX
PR 30-SEP-1999; 99KR-0042119.
XX
PA (CHOI/) CHOI E J.
XX
PI Choi EJ, Han PR, Kim IJ, Park BY, Park JH;
XX
DR WPI; 2001-569558/64.
DR N-PSDB; AAI66505.
XX
PT Rat's skip (sapk interacting protein) selectively inhibiting
PT stress-activated protein kinase and gene thereof -
XX
PS Claim 2; Page 20-24; 34pp; Korean.
XX
CC The present invention provides the protein and coding sequences of
CC different versions of the rat stress-activated kinase (SAPK) interacting
CC protein (SKIP). This protein is capable of modulating cell death, and
CC can be used in the treatment of cancer, inflammation, apoplexy and
CC similar diseases. The present sequence is one version of the protein
CC of the invention.
XX
SQ Sequence 708 AA;

Query Match 93.5%; Score 3544.5; DB 22; Length 708;

```
  Best Local Similarity   94.0%;  Pred. No. 1.2e-268;
  Matches  669;  Conservative   11;  Mismatches   27;  Indels    5;  Gaps    4;

Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
           |||||| || |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAERES-GLSGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 59

Qy      61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
           |||||||||||||||||  | ||||||||||||||||  ||||||||||| | ||
Db      60 QCKDTLSLRPPRAGLLSAGSSGSAGSRLQAEMLQMDLIDAASDTPGAEDDEEDDD-ELAA 118

Qy     120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
           :||| || |||||||||| ||  ||| | ||||||||||||||||||||||||||||||
Db     119 QRPGVGPSKAESGQEPASRSQG--QGQGPGTGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 176

Qy     180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
           |||||||||||||||||||||||||||||||||||||:|||||  | || ||||||||||
Db     177 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSDELPPQGSPVPTQDRGTSTDSPCRR 236

Qy     240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
           : ||||||| |||| ||||||||||||||||||||||||||||||||| ||||| |
Db     237 TAATQMAPPSGPPATAPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDPAEPTSTF 296

Qy     300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
           |||||||||||||||||  |||||||||||:||||||||:|||||||||||||||||||
Db     297 LPPTESRMSVSSDPDPAAYSVTAGRPHPSISEEDEGFDCLSSPEQAEPPGGGWRGSLGEP 356

Qy     360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     357 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASA 416

Qy     420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
           ||||||||||||||||:|:|| ||||||||||||||||||||||||||||||||||||||
Db     417 SSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 476

Qy     480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
           |||:|||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     477 FSCVINGEEHEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 536

Qy     540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
           |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db     537 PAYYAIEVTKEPEHMAALAKNSDWIDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 596

Qy     600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
           |||||||||||||||||||||||::||||||||||||||||||||||||||||||||||||
Db     597 TVHFNPPSSCVLEISVRGVKIGVKADEAQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 656

Qy     660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     657 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 708


RESULT 3
AAG78708
ID   AAG78708 standard; Protein; 715 AA.
XX
AC   AAG78708;
XX
DT   11-DEC-2001  (first entry)
XX
DE   Rat SAPK interacting protein #3.
XX
KW   Rat; SKIP; SAPK interacting protein; stress-activated kinase;
KW   cancer; inflammation; apoplexy; cell death; apoptosis.
XX
OS   Rattus norvegicus.
XX
FH   Key             Location/Qualifiers
FT   Domain          133..290
FT                   /label= JNK_binding_domain
FT   Domain          373..377
FT                   /label= proline_rich_domain
FT   Domain          491..553
FT                   /label= SH3_domain
FT   Domain          536..601
FT                   /label= helix-loop-helix_motif
FT   Domain          570..703
FT                   /label= phosphotyrosine_interaction_domain
XX
PN   KR2001029353-A.
XX
PD   06-APR-2001.
XX
PF   30-SEP-1999;   99KR-0042119.
XX
PR   30-SEP-1999;   99KR-0042119.
XX
PA   (CHOI/) CHOI E J.
XX
PI   Choi EJ,  Han PR,  Kim IJ,  Park BY,  Park JH;
XX
DR   WPI; 2001-569558/64.
DR   N-PSDB; AAI66506.
XX
PT   Rat's skip (sapk interacting protein) selectively inhibiting
PT   stress-activated protein kinase and gene thereof -
XX
PS   Claim 2; Page 27-30; 34pp; Korean.
XX
CC   The present invention provides the protein and coding sequences of
CC   different versions of the rat stress-activated kinase (SAPK) interacting
CC   protein (SKIP). This protein is capable of modulating cell death, and
CC   can be used in the treatment of cancer, inflammation, apoplexy and
CC   similar diseases. The present sequence is one version of the protein
CC   of the invention.
XX
SQ   Sequence   715 AA;

  Query Match             93.5%;  Score 3544.5;  DB 22;  Length 715;
  Best Local Similarity   94.0%;  Pred. No. 1.2e-268;
  Matches  669;  Conservative   11;  Mismatches   27;  Indels    5;  Gaps    4;

Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
           |||||| || |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8 MAERES-GLSGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 66

Qy      61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
           |||||||||||||||||  | ||||||||||||||||  ||||||||||| | ||
Db      67 QCKDTLSLRPPRAGLLSAGSSGSAGSRLQAEMLQMDLIDAASDTPGAEDDEEDDD-ELAA 125

Qy     120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
           :||| || |||||||||| ||  ||| | ||||||||||||||||||||||||||||||
Db     126 QRPGVGPSKAESGQEPASRSQG--QGQGPGTGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 183

Qy     180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
           |||||||||||||||||||||||||||||||||||||:|||||  | || ||||||||||
Db     184 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSDELPPQGSPVPTQDRGTSTDSPCRR 243

Qy     240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
           : ||||||| |||| ||||||||||||||||||||||||||||||||| ||||| |
Db     244 TAATQMAPPSGPPATAPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDPAEPTSTF 303

Qy     300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
           |||||||||||||||||  |||||||||||:||||||||:|||||||||||||||||||
Db     304 LPPTESRMSVSSDPDPAAYSVTAGRPHPSISEEDEGFDCLSSPEQAEPPGGGWRGSLGEP 363

Qy     360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASA 423

Qy     420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
           ||||||||||||||||:|:|| ||||||||||||||||||||||||||||||||||||||
Db     424 SSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 483

Qy     480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
           |||:|||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     484 FSCVINGEEHEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 543

Qy     540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
           ||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db     544 PAYYAIEVTKEPEHMAALAKNSDWIDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 603

Qy     600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
           |||||||||||||||||||||||||::|||||||||||||||||||||||||||||||||||
Db     604 TVHFNPPSSCVLEISVRGVKIGVKADEAQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 663

Qy     660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     664 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 715


RESULT 4
AAG78825
ID   AAG78825 standard; Protein; 707 AA.
XX
AC   AAG78825;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Murine SKIP-1b.
XX
KW   Murine; SKIP-1b; SAPK-interacting protein; phosphorylase; JNK; SAPK;
KW   c-Jun N-terminal kinase; Stress-Activated Protein Kinase;
KW   cell death related disease.
XX
OS   Mus musculus.
XX
PN   KR2001029352-A.
XX
PD   06-APR-2001.
XX
PF   30-SEP-1999;   99KR-0042118.
XX
PR   30-SEP-1999;   99KR-0042118.
XX
PA   (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
XX
PI   Choi IY,  Choi UJ,  Han PR,  Lee GU,  Lee JG,  Lee SH;
XX
DR   WPI; 2001-563617/63.
DR   N-PSDB; AAI65090.
XX
PT   Mouse skip (SAPK-interacting protein) controlling activation of human
PT   cell phosphorylase, jnk, and its gene, useful for treating cell death
PT   related diseases -
XX
PS   Claim 2; Page 18-21; 37pp; Korean.
XX
CC   The present sequence is the protein sequence for murine SKIP-1b
CC   (SAPK-interacting protein), which selectively inhibits or controls the
CC   activation of a phosphorylase, JNK/SAPK (c-Jun N-terminal kinase/
CC   Stress-Activated Protein Kinase), activated by various kinds of stress.
CC   SKIP and its gene are useful in treating cell death related diseases.
XX
SQ   Sequence   707 AA;

  Query Match             93.1%;  Score 3529;  DB 22;  Length 707;
  Best Local Similarity   93.3%;  Pred. No. 1.9e-267;
  Matches  665;  Conservative   13;  Mismatches   27;  Indels    8;  Gaps    4;

Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAERES-GLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 59

Qy      61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
           |||||||||||||||||||  | ||||||||||||||| |||||||||||:::|:| ||
Db      60 QCKDTLSLRPPRAGLLSAGSSGSAGSRLQAEMLQMDLIDAAGDTPGAEDDEEEEDDELAA 119

Qy     120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
           :||| ||||||| |:|| | |||        ||||||||||||||||||||||||||||||
Db     120 QRPGVGPPKAESNQDPAPRSQGQG-----ATGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 174

Qy     180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
           |||||||||||||||||||||||||||||||||:|||||  | || ||||||||||||
Db     175 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSDELPPQGSPVPTQDRGTSTDSPCRR 234

Qy     240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
           | ||||||| ||||  ||||||||||||||||||||||||||||||||||||| ||||| |
Db     235 SAATQMAPPSGPPATAPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDPAEPTSTF 294

Qy     300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
           :|||||||||||||||||  ||||||||||:|||||||||||||||||||||||||||||
Db     295 MPPTESRMSVSSDPDPAAYSVTAGRPHPSISEEDEGFDCLSSPERAEPPGGGWRGSLGEP 354

Qy     360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     355 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASA 414

Qy     420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
           |||||||||||||||:t:|| ||||||||||||||||||||||||||||||||||||||||
Db     415 SSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 474

Qy     480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
           |||:||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 FSCVINGEEHEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 534

Qy     540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTR-R 598
           |||||||||||||||||||||||:|||||||||||||||||||||||||||||||||| |
Db     535 PAYYAIEVTKEPEHMAALAKNSDWIDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRPR 594

Qy     599 LTVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFG 658
           |||||||||||||||||||||||||: ||||||||||||||||||||||||||||||||
Db     595 LTVHFNPPSSCVLEISVRGVKIGVKADDALEAKGNKCSHFFQLKNISFCGYHPKNNKYFG 654

Qy     659 FITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     655 FITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 707


RESULT 5
AAW80601
ID   AAW80601 standard; Protein; 714 AA.
XX
AC   AAW80601;
XX
DT   15-FEB-1999  (first entry)
XX
DE   Rat transcription factor islet-brain 1 (IB1).
XX
KW   IB1; islet-brain 1; transcription factor; rat; diabetes;
KW   dementia; Parkinson's disease; Alzheimer's disease; epilepsy;
KW   neuroblastoma; glioblastoma; apoptosis; cancer; autoimmune disease;
KW   systemic lupus erythematosus; myocardial infarction; ischaemia;
KW   diagnosis; therapy.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   Region          31..61
FT                   /note= "alpha-helical structure"
FT   Region          114..125
FT                   /note= "alpha-helical structure"
FT   Region          292..366
FT                   /note= "proline-rich region"
FT   Peptide         163..190
FT                   /note= "putative nuclear localisation signal"
FT   Peptide         242..270
FT                   /note= "putative nuclear localisation signal"
XX
PN   WO9844106-A1.
XX
PD   08-OCT-1998.
```

```
XX
PF   02-APR-1998;   98WO-GB00972.
XX
PR   15-MAY-1997;   97GB-0009920.
PR   03-APR-1997;   97GB-0006731.
XX
PA   (KIDD/) KIDDLE S J.
PA   (NICO/) NICOD P.
PA   (WAEB/) WAEBER G.
XX
PI   Bonny C,  Waeber G;
XX
DR   WPI; 1998-568278/48.
DR   N-PSDB; AAV62462.
XX
PT   New isolated transcription factor islet-brain 1 - used to develop
PT   products for treating e.g. diabetes, neurodegenerative disorders,
PT   cancers, autoimmune disease, heart disease or epilepsy
XX
PS   Claim 1; Fig 1A; 111pp; English.
XX
CC   This is the amino acid sequence of rat islet-brain 1 (IB1), a novel
CC   transcriptional activator that is involved in the control of the
CC   GLUT2 and insulin genes by interacting with homologous
CC   cis-regulatory elements of the GLUT2 and insulin gene promoters.
CC   The amino acid sequence was deduced from the sequence of isolated
CC   IB1 cDNA (see AAV80601) derived from rat insulinoma INS-1 cells.
CC   Human IB1 polynucleotide (see AAV62463) and polypeptide (see AAW80602)
CC   are also claimed.  IB1 polypeptides, nucleic acids, agonists and
CC   antagonists can be used in the treatment or diagnosis of diabetes,
CC   neurodegenerative disorders or cancers, to inhibit cells from
CC   undergoing apoptosis, to treat dementia, Parkinsonism, Alzheimer's
CC   disease, neuronal disabilities such as speech disorders and memory
CC   alteration, autoimmune diseases affecting the CNS such as systemic
CC   lupus erythematosus, diabetes, heart diseases such as myocardial
CC   infarct and ischemia or brain attack, neurological disorders, e.g.
CC   neuroblastoma, glioblastoma or cancers, or to promote apoptosis in
CC   cells, or treat refractory epilepsy.  They can also be used for
CC   screening test substances for IB1 biological activity.  IB1
CC   producing cells can be used therapeutically to produce IB1 in a
CC   subject.
XX
SQ   Sequence   714 AA;

  Query Match             92.4%;  Score 3503;  DB 19;  Length 714;
  Best Local Similarity   93.3%;  Pred. No. 2.1e-265;
  Matches  664;  Conservative   11;  Mismatches   31;  Indels    6;  Gaps    5;

Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
           |||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8 MAERES-GLSGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 66

Qy      61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
           |||||||||||||||||||  | ||||||||||||||||  ||||||||||||| | ||
Db      67 QCKDTLSLRPPRAGLLSAGSSGSAGSRLQAEMLQMDLIDAASDTPGAEDDEEDDD-ELAA 125

Qy     120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
           :||| || |||||||||||| ||  |||  | | ||||||||||||||||||||||||||
Db     126 QRPGVGPSKAESGQEPASRSQG--QGQGPGTGCGDTYRPKRPTTLNLFPQVPRSQDTLNN 183

Qy     180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
           ||||||||||||||||||||||||||||||||:|||||  | || ||||||||||||
Db     184 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSDELPPQGSPVPTQDRGTSTDSPCRR 243

Qy     240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
           : ||||||| ||||  ||||||||||||    ||||||||||||||||||||| ||||| |
Db     244 TAATQMAPPSGPPATAPGGRGHSHRDR-SISADVRLEATEEIYLTPVQRPPDPAEPTSTF 302

Qy     300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
           |||||||||||||||||  ||||||||||||:||||||||||:||||||||||||||||
Db     303 LPPTESRMSVSSDPDPAAYSVTAGRPHPSISEEDEGFDCLSSPEQAEPPGGGWRGSLGEP 362

Qy     360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     363 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASA 422

Qy     420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
           |||||||||||||||:|:|| ||||||||||||||||||||||||||||||||||||||||
Db     423 SSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 482

Qy     480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
           |||:||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     483 FSCVINGEEHEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 542

Qy     540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
           |||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||
Db     543 PAYYAIEVTKEPEHMAALAKNSDWIDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 602

Qy     600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
           |||||||||||||||||||||||||::|||||||||||||||||||||||||||||||||
Db     603 TVHFNPPSSCVLEISVRGVKIGVKADEAQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 662

Qy     660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     663 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 714


RESULT 6
AAG78706
ID   AAG78706 standard; Protein; 699 AA.
XX
AC   AAG78706;
XX
DT   11-DEC-2001  (first entry)
XX
DE   Rat SAPK interacting protein #1.
XX
KW   Rat; SKIP; SAPK interacting protein; stress-activated kinase;
KW   cancer; inflammation; apoplexy; cell death; apoptosis.
XX
OS   Rattus norvegicus.
XX
FH   Key             Location/Qualifiers
FT   Domain          117..274
FT                   /label= JNK_binding_domain
FT   Domain          357..361
FT                   /label= proline_rich_domain
FT   Domain          475..537
FT                   /label= SH3_domain
FT   Domain          520..585
FT                   /label= helix-loop-helix_motif
FT   Domain          554..688
FT                   /label= phosphotyrosine_interaction_domain
XX
PN   KR2001029353-A.
XX
PD   06-APR-2001.
XX
PF   30-SEP-1999;   99KR-0042119.
XX
PR   30-SEP-1999;   99KR-0042119.
XX
PA   (CHOI/) CHOI E J.
XX
PI   Choi EJ,  Han PR,  Kim IJ,  Park BY,  Park JH;
XX
DR   WPI; 2001-569558/64.
DR   N-PSDB; AAI66504.
XX
PT   Rat's skip (sapk interacting protein) selectively inhibiting
PT   stress-activated protein kinase and gene thereof -
XX
PS   Claim 2; Page 14-17; 34pp; Korean.
XX
CC   The present invention provides the protein and coding sequences of
CC   different versions of the rat stress-activated kinase (SAPK) interacting
```

```
CC   protein (SKIP). This protein is capable of modulating cell death, and
CC   can be used in the treatment of cancer, inflammation, apoplexy and
CC   similar diseases. The present sequence is one version of the protein
CC   of the invention.
XX
SQ   Sequence   699 AA;

  Query Match             89.4%;  Score 3389;  DB 22;  Length 699;
  Best Local Similarity   94.0%;  Pred. No. 1.7e-256;
  Matches  637;  Conservative   11;  Mismatches   26;  Indels    4;  Gaps    3;

Qy      35 LTHDISLEEFEDEDLSEITDECGISLQCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQ 93
           |||||||||||||||||||||||||||||||||||||||||||||  | ||||||||||||
Db      25 LTHDISLEEFEDEDLSEITDECGISLQCKDTLSLRPPRAGLLSAGSSGSAGSRLQAEMLQ 84

Qy      94 MDLIDATGDTPGAEDDEEDDDEERAARRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSG 153
           ||||||  |||||||||||||| | ||:||| || |||||||||| ||  |||  | |||
Db      85 MDLIDAASDTPGAEDDEEDDD-ELAAQRPGVGPSKAESGQEPASRSQG--QGQGPGTGSG 141

Qy     154 DTYRPKRPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICL 213
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142 DTYRPKRPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICL 201

Qy     214 SEELPPQSGPAPTTDRGTSTDSPCRRSTATQMAPPGGPPAAPPGGRGHSHRDRIHYQADV 273
           |:|||||  | || ||||||||||||: ||||||| |||| ||||||||||||||||||
Db     202 SDELPPQGSPVPTQDRGTSTDSPCRRTAATQMAPPSGPPATAPGGRGHSHRDRIHYQADV 261

Qy     274 RLEATEEIYLTPVQRPPDAAEPTSAFLPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEE 333
           |||||||||||||||| ||||| |||||||||||||||||  ||||||||||:
Db     262 RLEATEEIYLTPVQRPPDPAEPTSTFLPPTESRMSVSSDPDPAAYSVTAGRPHPSISEED 321

Qy     334 EGFDCLSSPERAEPPGGGWRGSLGEPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQLEL 393
           ||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db     322 EGFDCLSSPEQAEPPGGGWRGSLGEPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQLEL 381

Qy     394 VSLRPCFGDYSDESDSATVYDNCASVSSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPD 453
           |||||||||||||||||||||||| ||||||||||||||:|:|| ||||||||||||
Db     382 VSLRPCFGDYSDESDSATVYDNCASASSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPD 441

Qy     454 VHFSKKFLNVFMSGRSRSSSAESFGLFSCIINGEEQEQTHRAIFRFVPRHEDELELEVDD 513
           |||||||||||||||||||||||||:||||| ||||||||||||||||||||||||||||
Db     442 VHFSKKFLNVFMSGRSRSSSAESFGLFSCVINGEEHEQTHRAIFRFVPRHEDELELEVDD 501

Qy     514 PLLVELQAEDYWYEAYNMRTGARGVFPAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLG 573
           |||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||
Db     502 PLLVELQAEDYWYEAYNMRTGARGVFPAYYAIEVTKEPEHMAALAKNSDWIDQFRVKFLG 561

Qy     574 SVQVPYHKGNDVLCAAMQKIATTRRLTVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGN 633
           |||||||||||||||||||||||||||||||||||||||||||||||||||||::||||||
Db     562 SVQVPYHKGNDVLCAAMQKIATTRRLTVHFNPPSSCVLEISVRGVKIGVKADEAQEAKGN 621

Qy     634 KCSHFFQLKNISFCGYHPKNNKYFGFITKHPADHRFACHVFVSEDSTKALAESVGRAFQQ 693
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     622 KCSHFFQLKNISFCGYHPKNNKYFGFITKHPADHRFACHVFVSEDSTKALAESVGRAFQQ 681

Qy     694 FYKQFVEYTCPTEDIYLE 711
           ||||||||||||||||||
Db     682 FYKQFVEYTCPTEDIYLE 699


RESULT 7
AAG78826
ID   AAG78826 standard; Protein; 698 AA.
XX
AC   AAG78826;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Murine SKIP-2a.
XX
KW   Murine; SKIP-2a; SAPK-interacting protein; phosphorylase; JNK; SAPK;
KW   c-Jun N-terminal kinase; Stress-Activated Protein Kinase;
KW   cell death related disease.
XX
OS   Mus musculus.
XX
PN   KR2001029352-A.
XX
PD   06-APR-2001.
XX
PF   30-SEP-1999;   99KR-0042118.
XX
PR   30-SEP-1999;   99KR-0042118.
XX
PA   (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
XX
PI   Choi IY,  Choi UJ,  Han PR,  Lee GU,  Lee JG,  Lee SH;
XX
DR   WPI; 2001-563617/63.
DR   N-PSDB; AAI65091.
XX
PT   Mouse skip (SAPK-interacting protein) controlling activation of human
PT   cell phosphorylase, jnk, and its gene, useful for treating cell death
PT   related diseases -
XX
PS   Claim 2; Page 24-27; 37pp; Korean.
XX
CC   The present sequence is the protein sequence for murine SKIP-2a
CC   (SAPK-interacting protein), which selectively inhibits or controls the
CC   activation of a phosphorylase, JNK/SAPK (c-Jun N-terminal kinase/
CC   Stress-Activated Protein Kinase), activated by various kinds of stress.
CC   SKIP and its gene are useful in treating cell death related diseases.
XX
SQ   Sequence   698 AA;

  Query Match             88.8%;  Score 3367.5;  DB 22;  Length 698;
  Best Local Similarity   93.1%;  Pred. No. 8.2e-255;
  Matches  632;  Conservative   13;  Mismatches   27;  Indels    7;  Gaps    3;

Qy      35 LTHDISLEEFEDEDLSEITDECGISLQCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQ 93
           |||||||||||||||||||||||||||||||||||||||||||||  | ||||||||||||
Db      25 LTHDISLEEFEDEDLSEITDECGISLQCKDTLSLRPPRAGLLSAGSSGSAGSRLQAEMLQ 84

Qy      94 MDLIDATGDTPGAEDDEEDDDEERAARRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSG 153
           |||||| ||||||||||::|:| ||:||| ||||||| |:|| | |||        |||
Db      85 MDLIDAAGDTPGAEDDEEEEDDELAAQRPGVGPPKAESNQDPAPRSQGQG-----ATGSG 139

Qy     154 DTYRPKRPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICL 213
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     140 DTYRPKRPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICL 199

Qy     214 SEELPPQSGPAPTTDRGTSTDSPCRRSTATQMAPPGGPPAAPPGGRGHSHRDRIHYQADV 273
           |:|||||  | || ||||||||||||| |||||| |||| ||||||||||||||||||
Db     200 SDELPPQGSPVPTQDRGTSTDSPCRRSAATQMAPPSGPPATAPGGRGHSHRDRIHYQADV 259

Qy     274 RLEATEEIYLTPVQRPPDAAEPTSAFLPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEE 333
           |||||||||||||||| ||||| |:|||||||||||||||||  ||||||||||:
Db     260 RLEATEEIYLTPVQRPPDPAEPTSTFMPPTESRMSVSSDPDPAAYSVTAGRPHPSISEED 319

Qy     334 EGFDCLSSPERAEPPGGGWRGSLGEPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQLEL 393
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     320 EGFDCLSSPERAEPPGGGWRGSLGEPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQLEL 379

Qy     394 VSLRPCFGDYSDESDSATVYDNCASVSSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPD 453
           |||||||||||||||||||||||| ||||||||||||||:|:|| ||||||||||||
Db     380 VSLRPCFGDYSDESDSATVYDNCASASSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPD 439

Qy     454 VHFSKKFLNVFMSGRSRSSSAESFGLFSCIINGEEQEQTHRAIFRFVPRHEDELELEVDD 513
           |||||||||||||||||||||||||||:||||| ||||||||||||||||||||||||||||
Db     440 VHFSKKFLNVFMSGRSRSSSAESFGLFSCVINGEEHEQTHRAIFRFVPRHEDELELEVDD 499

Qy     514 PLLVELQAEDYWYEAYNMRTGARGVFPAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLG 573
           |||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||
Db     500 PLLVELQAEDYWYEAYNMRTGARGVFPAYYAIEVTKEPEHMAALAKNSDWIDQFRVKFLG 559

Qy     574 SVQVPYHKGNDVLCAAMQKIATTR-RLTVHFNPPSSCVLEISVRGVKIGVKADDSQEAKG 632
```

```
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||: ||||
Db    560 SVQVPYHKGNDVLCAAMQKIATTRPRLTVHFNPPSSCVLEISVRGVKIGVKADDALEAKG 619

Qy    633 NKCSHFFQLKNISFCGYHPKNNKYFGFITKHPADHRFACHVFVSEDSTKALAESVGRAFQ 692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    620 NKCSHFFQLKNISFCGYHPKNNKYFGFITKHPADHRFACHVFVSEDSTKALAESVGRAFQ 679

Qy    693 QFYKQFVEYTCPTEDIYLE 711
          |||||||||||||||||||
Db    680 QFYKQFVEYTCPTEDIYLE 698


RESULT 8
AAW81525
ID   AAW81525 standard; Protein; 660 AA.
XX
AC   AAW81525;
XX
DT   01-MAR-1999  (first entry)
XX
DE   Murine JNK-interacting protein 1 (JIP-1).
XX
KW   JIP-1; JNK-interacting protein; c-Jun NH2-terminal kinase;
KW   signal transduction; inhibitor; mouse; neurodegenerative disease;
KW   Parkinson's disease; Alzheimer's disease; blood clot; stroke;
KW   malignancy; cancer; leukaemia; autoimmune disease; inflammation;
KW   apoptosis; therapy; diagnosis.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   Domain          127..281
FT                   /label= JBD
FT                   /note= "claimed JNK binding domain"
FT   Domain          148..174
FT                   /note= "claimed JBD core"
FT   Domain          491..540
FT                   /label= SH3
XX
PN   WO9849188-A1.
XX
PD   05-NOV-1998.
XX
PF   28-APR-1998;   98WO-US08513.
XX
PR   28-APR-1997;   97US-0819177.
XX
PA   (UYMA-) UNIV MASSACHUSETTS.
XX
PI   Davis RJ,  Dickens M;
XX
DR   WPI; 1999-024042/02.
DR   N-PSDB; AAV69289.
XX
PT   c-Jun NH2-terminal kinase (JNK)-interacting protein 1 - used to
PT   treat neurodegenerative disease, blood clot, leukaemia, autoimmune
PT   disease, and inflammation
XX
PS   Claim 5; Fig 1B; 95pp; English.
XX
CC   This is the amino acid sequence of murine JNK-interacting protein 1
CC   (JIP-1), a novel cytoplasmic anchor protein that specifically binds
CC   to and inhibits the biological effects of JNK (c-Jun NH2-terminal
CC   kinase), including the initiation of apoptosis and oncogenic
CC   transformation.  The sequence is predicted from a full-length cDNA
CC   clone (see AAV69289) isolated from a mouse brain cDNA lkibrary.  The
CC   invention provides JIP-1 nucleic acids and polypeptides (see also
CC   AAW81535-45), expression vectors and host cells.  The JIP-1
CC   polypeptides and nucleic acids (including antisense and ribozymes)
CC   can be used in the manufacture of a medicament for treating a
CC   pathological condition associated with abnormal expression or
CC   activity of JNK, such as a neurodegenerative disease (selected from
CC   Parkinson's disease and Alzheimer's disease), a blood clot, stroke,
CC   malignancy, leukaemia, an autoimmune disease or inflammation (all
CC   claimed).
XX
SQ   Sequence   660 AA;

  Query Match             86.0%;  Score 3262.5;  DB 20;  Length 660;
  Best Local Similarity   87.1%;  Pred. No. 1.3e-246;
  Matches  620;  Conservative   13;  Mismatches   26;  Indels   53;  Gaps    4;

Qy      1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MAERES-GLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 59

Qy     61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
          |||||||||||||||||||  | ||||||||||||||:|| |||||||||||::|:| ||
Db     60 QCKDTLSLRPPRAGLLSAGSSGSAGSRLQAEMLQMDLVDAAGDTPGAEDDEEEEDDELAA 119

Qy    120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
          :||| ||||||| |:|| |    |||| | ||||||||||||||||||||||||||||||
Db    120 QRPGVGPPKAESNQDPAPR----SQGQGPGTGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 175

Qy    180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
          ||||||||||||||||||||||||||||||||||:||||  | || ||||||||||||
Db    176 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSDELPPQGSPVPTQDRGTSTDSPCRR 235

Qy    240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
          | ||||||| |||| ||||||||||||||||||||||||||||||||||||| ||||| |
Db    236 SAATQMAPPSGPPATAPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDPAEPTSTF 295

Qy    300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
          :|||||||||||||||||  ||||||||||:|||||||||||||||||||||||||||||
Db    296 MPPTESRMSVSSDPDPAAYSVTAGRPHPSISEEDEGFDCLSSPERAEPPGGGWRGSLGEP 355

Qy    360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    356 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASA 415

Qy    420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
          |||||||||||||||:|:|| ||||||||||||||||||||||||||||||||||||||||
Db    416 SSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 475

Qy    480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
          |||:||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    476 FSCVINGEEHEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 535

Qy    540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
          ||||||||||||||||||
Db    536 PAYYAIEVTKEPEHMAALAKN---------------------------------------- 556

Qy    600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
                  |||||||||||||||||: |||||||||||||||||||||||||||||||||||
Db    557 --------SCVLEISVRGVKIGVKADDALEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 608

Qy    660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 660


RESULT 9
AAG78824
ID   AAG78824 standard; Protein; 673 AA.
XX
AC   AAG78824;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Murine SKIP-2b.
XX
KW   Murine; SKIP-2b; SAPK-interacting protein; phosphorylase; JNK; SAPK;
KW   c-Jun N-terminal kinase; Stress-Activated Protein Kinase;
KW   cell death related disease.
XX
OS   Mus musculus.
```

```
XX
PN   KR2001029352-A.
XX
PD   06-APR-2001.
XX
PF   30-SEP-1999;   99KR-0042118.
XX
PR   30-SEP-1999;   99KR-0042118.
XX
PA   (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
XX
PI   Choi IY,  Choi UJ,  Han PR,  Lee GU,  Lee JG,  Lee SH;
XX
DR   WPI; 2001-563617/63.
DR   N-PSDB; AAI65089.
XX
PT   Mouse skip (SAPK-interacting protein) controlling activation of human
PT   cell phosphorylase, jnk, and its gene, useful for treating cell death
PT   related diseases -
XX
PS   Claim 2; Page 12-15; 37pp; Korean.
XX
CC   The present sequence is the protein sequence for murine SKIP-2b
CC   (SAPK-interacting protein), which selectively inhibits or controls the
CC   activation of a phosphorylase, JNK/SAPK (c-Jun N-terminal kinase/
CC   Stress-Activated Protein Kinase), activated by various kinds of stress.
CC   SKIP and its gene are useful in treating cell death related diseases.
XX
SQ   Sequence   673 AA;

  Query Match             85.7%;  Score 3248;  DB 22;  Length 673;
  Best Local Similarity   90.0%;  Pred. No. 1.8e-245;
  Matches  610;  Conservative   13;  Mismatches   25;  Indels   30;  Gaps    3;

Qy      35 LTHDISLEEFEDEDLSEITDECGISLQCKDTLSLRPPRAGLLSAGGGGAGSRLQAEMLQM 94
           |||||||||||||||||||||||||||||||||||||||                    |
Db      25 LTHDISLEEFEDEDLSEITDECGISLQCKDTLSLR-----------------------M 60

Qy      95 DLIDATGDTPGAEDDEEDDDEERAARRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGD 154
           ||||| |||||||||||::|:| ||:||| ||||||| |:|| | |||         ||||
Db      61 DLIDAAGDTPGAEDDEEEEDDELAAQRPGVGPPKAESNQDPAPRSQGQG-----ATGSGD 115

Qy     155 TYRPKRPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLS 214
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116 TYRPKRPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLS 175

Qy     215 EELPPQSGPAPTTDRGTSTDSPCRRSTATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVR 274
           :||||  | || |||||||||||| ||||||| |||  |||||||||||||||||||||||
Db     176 DELPPQGSPVPTQDRGTSTDSPCRRSAATQMAPPSGPPATAPGGRGHSHRDRIHYQADVR 235

Qy     275 LEATEEIYLTPVQRPPDAAEPTSAFLPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEE 334
           |||||||||||||||| ||||| |:|||||||||||||||||  ||||||||||||:|
Db     236 LEATEEIYLTPVQRPPDPAEPTSTFMPPTESRMSVSSDPDPAAYSVTAGRPHPSISEEDE 295

Qy     335 GFDCLSSPERAEPPGGGWRGSLGEPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELV 394
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     296 GFDCLSSPERAEPPGGGWRGSLGEPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELV 355

Qy     395 SLRPCFGDYSDESDSATVYDNCASVSSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDV 454
           |||||||||||||||||||||| |||||||||||||||:|:|| ||||||||||||||
Db     356 SLRPCFGDYSDESDSATVYDNCASASSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPDV 415

Qy     455 HFSKKFLNVFMSGRSRSSSAESFGLFSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDP 514
           |||||||||||||||||||||||||||:||||| ||||||||||||||||||||||||||||
Db     416 HFSKKFLNVFMSGRSRSSSAESFGLFSCVINGEEHEQTHRAIFRFVPRHEDELELEVDDP 475

Qy     515 LLVELQAEDYWYEAYNMRTGARGVFPAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGS 574
           ||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db     476 LLVELQAEDYWYEAYNMRTGARGVFPAYYAIEVTKEPEHMAALAKNSDWIDQFRVKFLGS 535

Qy     575 VQVPYHKGNDVLCAAMQKIATTR-RLTVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGN 633
           |||||||||||||||||||| |||||||||||||||||||||||||: |||||
Db     536 VQVPYHKGNDVLCAAMQKIATTRPRLTVHFNPPSSCVLEISVRGVKIGVKADDALEAKGN 595

Qy     634 KCSHFFQLKNISFCGYHPKNNKYFGFITKHPADHRFACHVFVSEDSTKALAESVGRAFQQ 693
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     596 KCSHFFQLKNISFCGYHPKNNKYFGFITKHPADHRFACHVFVSEDSTKALAESVGRAFQQ 655

Qy     694 FYKQFVEYTCPTEDIYLE 711
           ||||||||||||||||||
Db     656 FYKQFVEYTCPTEDIYLE 673


RESULT 10
ABB04788
ID   ABB04788 standard; Protein; 659 AA.
XX
AC   ABB04788;
XX
DT   13-MAR-2002  (first entry)
XX
DE   LDL receptor binding protein JIP-1 SEQ ID NO:11.
XX
KW   Low density lipoprotein receptor binding protein; signal transduction;
KW   LDL receptor binding protein; LDL receptor signalling pathway.
XX
OS   Synthetic.
XX
PN   WO200184159-A2.
XX
PD   08-NOV-2001.
XX
PF   24-APR-2001; 2001WO-US13214.
XX
PR   01-MAY-2000; 2000US-0562737.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Herz J,  Gotthardt M;
XX
DR   WPI; 2002-082855/11.
XX
PT   Detecting stress that alters interaction of LDL receptor binding
PT   polypeptide with LDL receptor interaction domain, comprises detecting
PT   difference in stress-biased and unbiased interaction of peptide and
PT   domain in a system  -
XX
PS   Disclosure; Page 34-36; 200pp; English.
XX
CC   The present invention describes a method for detecting a stress that
CC   alters a functional interaction of a low density lipoprotein (LDL)
CC   receptor binding protein (I) with an LDL receptor interaction domain
CC   (II). The method involves introducing a predetermined stress into a
CC   system which provides a stress-biased physical interaction of (I) with
CC   (II), where in the absence of the stress, the system provides an
CC   unbiased interaction of (I) and (II), and detecting the stress-biased
CC   interaction of (I) and (II), where a difference between BI and UI
CC   indicates that the stress alters the interaction of (I) and (II). (I)
CC   is selected from SEMCAP-1, JIP-1, PSD-95, JIP-2, Talin, OMP25, CAPON,
CC   PIP4,5 Kinase, Na channel brain 3, Mint1, ICAP-1 and APC subunit 10.
CC   The method is useful for detecting a stress that alters functional
CC   interaction of LDL receptor binding polypeptide with LDL receptor
CC   interaction domain. The method is useful for detecting and modulating
CC   signal transduction through LDL receptors. ABB04778 to ABB04909
CC   represent LDL receptor binding proteins which are used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   659 AA;

  Query Match             85.6%;  Score 3245;  DB 23;  Length 659;
  Best Local Similarity   87.1%;  Pred. No. 2.9e-245;
  Matches  620;  Conservative   12;  Mismatches   26;  Indels   54;  Gaps    5;

Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAERES-GLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 59
```

```
Qy      61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
           ||||||||||||||||||  | ||||||||||||||| |||||||||||::|:| ||
Db      60 QCKDTLSLRPPRAGLLSAGSSGSAGSRLQAEMLQMDLIDAAGDTPGAEDDEEEEDDELAA 119

Qy     120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
           :||| ||||||| |:|| |    |||| | ||||||||||||||||||||||||||||||
Db     120 QRPGVGPPKAESNQDPAPR----SQGQGPGTGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 175

Qy     180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
           ||||||||||||||||||||||||||||||||:|||||  | || ||||||||||||
Db     176 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSDELPPQGSPVPTQDRGTSTDSPCRR 235

Qy     240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
           | ||||||| ||||  ||||||||||||||||||||||||||||||||||||| ||||| |
Db     236 SAATQMAPPSGPPATAPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDPAEPTSTF 295

Qy     300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
           :|||||||||||||||||  ||||||||||||:|||||||||||||||||||||||||||
Db     296 MPPTESRMSVSSDPDPAAYSVTAGRPHPSISEEDEGFDCLSSPERAEPPGGGWRGSLGEP 355

Qy     360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     356 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASA 415

Qy     420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
           |||||||||||||||:|:|| ||||||||||||| ||||||||||||||||||||||||
Db     416 SSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPDV-FSKKFLNVFMSGRSRSSSAESFGL 474

Qy     480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
           |||:||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 FSCVINGEEHEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 534

Qy     540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
           ||||||||||||||||||||
Db     535 PAYYAIEVTKEPEHMAALAKN---------------------------------------- 555

Qy     600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
                   ||||||||||||||||||: ||||||||||||||||||||||||||||||||
Db     556 --------SCVLEISVRGVKIGVKADDALEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 607

Qy     660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     608 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 659
```

RESULT 11
AAG78827
ID AAG78827 standard; Protein; 617 AA.
XX
AC AAG78827;
XX
DT 18-DEC-2001 (first entry)
XX
DE Murine SKIP-3.
XX
KW Murine; SKIP-3; SAPK-interacting protein; phosphorylase; JNK; SAPK;
KW c-Jun N-terminal kinase; Stress-Activated Protein Kinase;
KW cell death related disease.
XX
OS Mus musculus.
XX
PN KR2001029352-A.
XX
PD 06-APR-2001.
XX
PF 30-SEP-1999; 99KR-0042118.
XX
PR 30-SEP-1999; 99KR-0042118.
XX
PA (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
XX
PI Choi IY, Choi UJ, Han PR, Lee GU, Lee JG, Lee SH;
XX
DR WPI; 2001-563617/63.
DR N-PSDB; AAI65092.
XX
PT Mouse skip (SAPK-interacting protein) controlling activation of human
PT cell phosphorylase, jnk, and its gene, useful for treating cell death
PT related diseases -
XX
PS Claim 2; Page 30-33; 37pp; Korean.
XX
CC The present sequence is the protein sequence for murine SKIP-3
CC (SAPK-interacting protein), which selectively inhibits or controls the
CC activation of a phosphorylase, JNK/SAPK (c-Jun N-terminal kinase/
CC Stress-Activated Protein Kinase), activated by various kinds of stress.
CC SKIP and its gene are useful in treating cell death related diseases.
XX
SQ Sequence 617 AA;

Query Match 81.9%; Score 3104; DB 22; Length 617;
Best Local Similarity 92.9%; Pred. No. 2.9e-234;
Matches 578; Conservative 13; Mismatches 25; Indels 6; Gaps 2;

```
Qy      91 MLQMDLIDATGDTPGAEDDEEDDDEERAARRPGAGPPKAESGQEPASRGQGQSQGQSQGP 150
           ||||||||| |||||||||||::|:| ||:||| ||||||| |:|| | |||
Db       1 MLQMDLIDAAGDTPGAEDDEEEEDDELAAQRPGVGPPKAESNQDPAPRSQGQG-----AT 55

Qy     151 GSGDTYRPKRPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEH 210
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56 GSGDTYRPKRPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEH 115

Qy     211 ICLSEELPPQSGPAPTTDRGTSTDSPCRRSTATQMAPPGGPPAAPPGGRGHSHRDRIHYQ 270
           ||||:||||| | || |||||||||||||| ||||||| |||| ||||||||||||||||
Db     116 ICLSDELPPQGSPVPTQDRGTSTDSPCRRSAATQMAPPSGPPATAPGGRGHSHRDRIHYQ 175

Qy     271 ADVRLEATEEIYLTPVQRPPDAAEPTSAFLPPTESRMSVSSDPDPAAYPSTAGRPHPSIS 330
           |||||||||||||||||||| ||||| |:|||||||||||||||||  ||||||||||
Db     176 ADVRLEATEEIYLTPVQRPPDPAEPTSTFMPPTESRMSVSSDPDPAAYSVTAGRPHPSIS 235

Qy     331 EEEEGFDCLSSPERAEPPGGGWRGSLGEPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQ 390
           ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     236 EEDEGFDCLSSPERAEPPGGGWRGSLGEPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQ 295

Qy     391 LELVSLRPCFGDYSDESDSATVYDNCASVSSPYESAIGEEYEEAPRPQPPACLSEDSTPD 450
           |||||||||||||||||||||||||||| ||||||||||||||||:|:|| ||||||||||
Db     296 LELVSLRPCFGDYSDESDSATVYDNCASASSPYESAIGEEYEEAPQPRPPTCLSEDSTPD 355

Qy     451 EPDVHFSKKFLNVFMSGRSRSSSAESFGLFSCIINGEEQEQTHRAIFRFVPRHEDELELE 510
           |||||||||||||||||||||||||||||:||||| ||||||||||||||||||||||||
Db     356 EPDVHFSKKFLNVFMSGRSRSSSAESFGLFSCVINGEEHEQTHRAIFRFVPRHEDELELE 415

Qy     511 VDDPLLVELQAEDYWYEAYNMRTGARGVFPAYYAIEVTKEPEHMAALAKNSDWVDQFRVK 570
           ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db     416 VDDPLLVELQAEDYWYEAYNMRTGARGVFPAYYAIEVTKEPEHMAALAKNSDWIDQFRVK 475

Qy     571 FLGSVQVPYHKGNDVLCAAMQKIATTR-RLTVHFNPPSSCVLEISVRGVKIGVKADDSQE 629
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||: |
Db     476 FLGSVQVPYHKGNDVLCAAMQKIATTRPRLTVHFNPPSSCVLEISVRGVKIGVKADDALE 535

Qy     630 AKGNKCSHFFQLKNISFCGYHPKNNKYFGFITKHPADHRFACHVFVSEDSTKALAESVGR 689
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     536 AKGNKCSHFFQLKNISFCGYHPKNNKYFGFITKHPADHRFACHVFVSEDSTKALAESVGR 595

Qy     690 AFQQFYKQFVEYTCPTEDIYLE 711
           ||||||||||||||||||||||
Db     596 AFQQFYKQFVEYTCPTEDIYLE 617
```

RESULT 12
ABB04797
ID ABB04797 standard; Protein; 659 AA.
XX
AC ABB04797;
XX

DT 13-MAR-2002 (first entry)
XX
DE LDL receptor binding protein JIP-1 SEQ ID NO:20.
XX
KW Low density lipoprotein receptor binding protein; signal transduction;
KW LDL receptor binding protein; LDL receptor signalling pathway.
XX
OS Synthetic.
XX
PN WO200184159-A2.
XX
PD 08-NOV-2001.
XX
PF 24-APR-2001; 2001WO-US13214.
XX
PR 01-MAY-2000; 2000US-0562737.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Herz J, Gotthardt M;
XX
DR WPI; 2002-082855/11.
XX
PT Detecting stress that alters interaction of LDL receptor binding
PT polypeptide with LDL receptor interaction domain, comprises detecting
PT difference in stress-biased and unbiased interaction of peptide and
PT domain in a system -
XX
PS Disclosure; Page 49-51; 200pp; English.
XX
CC The present invention describes a method for detecting a stress that
CC alters a functional interaction of a low density lipoprotein (LDL)
CC receptor binding protein (I) with an LDL receptor interaction domain
CC (II). The method involves introducing a predetermined stress into a
CC system which provides a stress-biased physical interaction of (I) with
CC (II), where in the absence of the stress, the system provides an
CC unbiased interaction of (I) and (II), and detecting the stress-biased
CC interaction of (I) and (II), where a difference between BI and UI
CC indicates that the stress alters the interaction of (I) and (II). (I)
CC is selected from SEMCAP-1, JIP-1, PSD-95, JIP-2, Talin, OMP25, CAPON,
CC PIP4,5 Kinase, Na channel brain 3, Mint1, ICAP-1 and APC subunit 10.
CC The method is useful for detecting a stress that alters functional
CC interaction of LDL receptor binding polypeptide with LDL receptor
CC interaction domain. The method is useful for detecting and modulating
CC signal transduction through LDL receptors. ABB04778 to ABB04909
CC represent LDL receptor binding proteins which are used in the
CC exemplification of the present invention.
XX
SQ Sequence 659 AA;

Query Match 75.8%; Score 2875; DB 23; Length 659;
Best Local Similarity 78.8%; Pred. No. 2.7e-216;
Matches 561; Conservative 18; Mismatches 79; Indels 54; Gaps 5;

```
Qy      1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
          |||||| ||| |||||||| |||||||| |||||||| |||||||| |||||||| |||||||||
Db      1 MAERES-GLGAGAASPPAASDFLGLHIASPENFRLTHDISFEEFEDEDLSGITDECGISL 59

Qy     61 QCKDTLSLRPPRAGLLSAGG-GGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
           ||||||||| |||||||  | ||||||| |||||||| |||||||| |::|:| ||
Db     60 HCKDTLSLRPIRAGLLSAGSKGSAGSRLQALMLQMDLIDAMGDTPGAEDDNEEEDDELAA 119

Qy    120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
          |||| ||||| | |:|| |    | ||  | ||| |||||||| |||||||| |||||
Db    120 RRPGVGPPKASSNQDPAPR----STGQGPGTGSGVTYRPKRPTTWNLFPQVPRSYDTLNN 175

Qy    180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
          |||| ||||||||| |||||||||:||||||||| :|||||  | || |||||| |||||
Db    176 NSLGAKHSWQDRVSDSSSPLKTGEETPPHEHICLFDELPPQGSPGPTQDRGTSTHSPCRR 235

Qy    240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
          | || |||| ||||  |||||||| |||||||| |||||||| ||||||| |:||| |
Db    236 SAATIMAPPSGPPAKAPGGRGHSHLDRIHYQADVMLEATEEIYLNPVQRPPDPAQPTSTF 295

Qy    300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
          :||| |||||||||| |||| ||| ||||||||: |||||||| |||||||| |||||
Db    296 MPPTRSRMSVSSDPSPAAYSVTAGTPHPSISEEDVGFDCLSSPEWAEPPGGGWRYSLGEP 355

Qy    360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
          |||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||
Db    356 PPPPAASLSSDTSADSYDSVKYTLEVDEHAQLELFSLRPCFGDYGDESDSATVYHNCASA 415

Qy    420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
          |||| |||||||||| |:|:|:| || |||||||||: |||||||| |||||||||:||||
Db    416 SSPYISAIGEEYEEKPQPRPPTCLLEDSTPDEPDM-FSKKFLNVFNSGRSRSSSAQSFGL 474

Qy    480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
          |||:| ||| ||||||:||||||||| |||||||||:|||||||||:||||||||| ||||
Db    475 FSCVIRGEEHEQTHRSIFRFVPRHETELELEVDDPVLVELQAEDYYYEAYNMRTGDRGVF 534

Qy    540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
          ||||| ||||||||| |||||
Db    535 PAYYAEEVTKEPEHMFALAKN---------------------------------------- 555

Qy    600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
                    |||| ||||||||| |||||||| |||||||| |||||||
Db    556 --------SCVLGISVRGVKIGHKADDALEAKINKCSHFFQLLNISFCGYHPMNNKYFGF 607

Qy    660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          || ||||||||| |||||||| |||||||| |||||||| |||||||| |||||||||
Db    608 ITNHPADHRFACQVFVSEDSTKRLAESVGRAFSQFYKQFVEYVCPTEDIYLE 659
```

RESULT 13
ABB04793
ID ABB04793 standard; Protein; 659 AA.
XX
AC ABB04793;
XX
DT 13-MAR-2002 (first entry)
XX
DE LDL receptor binding protein JIP-1 SEQ ID NO:16.
XX
KW Low density lipoprotein receptor binding protein; signal transduction;
KW LDL receptor binding protein; LDL receptor signalling pathway.
XX
OS Synthetic.
XX
PN WO200184159-A2.
XX
PD 08-NOV-2001.
XX
PF 24-APR-2001; 2001WO-US13214.
XX
PR 01-MAY-2000; 2000US-0562737.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Herz J, Gotthardt M;
XX
DR WPI; 2002-082855/11.
XX
PT Detecting stress that alters interaction of LDL receptor binding
PT polypeptide with LDL receptor interaction domain, comprises detecting
PT difference in stress-biased and unbiased interaction of peptide and
PT domain in a system -
XX
PS Disclosure; Page 43-44; 200pp; English.
XX
CC The present invention describes a method for detecting a stress that
CC alters a functional interaction of a low density lipoprotein (LDL)
CC receptor binding protein (I) with an LDL receptor interaction domain
CC (II). The method involves introducing a predetermined stress into a
CC system which provides a stress-biased physical interaction of (I) with
CC (II), where in the absence of the stress, the system provides an
CC unbiased interaction of (I) and (II), and detecting the stress-biased

```
CC   interaction of (I) and (II), where a difference between BI and UI
CC   indicates that the stress alters the interaction of (I) and (II). (I)
CC   is selected from SEMCAP-1, JIP-1, PSD-95, JIP-2, Talin, OMP25, CAPON,
CC   PIP4,5 Kinase, Na channel brain 3, Mint1, ICAP-1 and APC subunit 10.
CC   The method is useful for detecting a stress that alters functional
CC   interaction of LDL receptor binding polypeptide with LDL receptor
CC   interaction domain. The method is useful for detecting and modulating
CC   signal transduction through LDL receptors. ABB04778 to ABB04909
CC   represent LDL receptor binding proteins which are used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   659 AA;

  Query Match             75.5%;  Score 2864;  DB 23;  Length 659;
  Best Local Similarity   78.5%;  Pred. No. 1.9e-215;
  Matches  559;  Conservative   23;  Mismatches   76;  Indels   54;  Gaps    5;

Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
Db       1 MAEREA-GLGGGAASPDAASPFLGLHEASPPNFRLTFDISLEEFEDGDLSEITDECHISL 59

Qy      61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
Db      60 QCKDTLILRPPRAGLLKAGSSGSAGSLLQAEMLQMDMIDAAGDTPGNEDDEEEEDDQLAA 119

Qy     120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
Db     120 QRPGVGRPKAESNQDPSPR----SQGQGPGVGSGDTYRPKWPTTLNLFPQYPRSQDTLNN 175

Qy     180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
Db     176 ASLGKKHSWQDRVSRSSSPLETGEQTPPHEFICLSDELPPGGSPVPTQDRHTSTDSPCRR 235

Qy     240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
Db     236 IAATQMAPPSKPPATAPGGRLHSHRDRIHYMADVRLEATENIYLTPVQRPQDPAEPTSTF 295

Qy     300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
Db     296 RPPTESRMSVTSDPDPAAYSWTAGRPHPSIYEEDEGFDCLASPERAEPPGDGWRGSLGEP 355

Qy     360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
Db     356 EPPPRASLSSFTSALSYDSVGYTLVVDEHAHLELVSLRPCIGDYSDESDSKTVYDNCASA 415

Qy     420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
Db     416 LSPYESAIGEMYEEAPQPRPNTCLSEDSTPQEPDV-FSKKFRNVFMSGRSRTSSAESFGL 474

Qy     480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
Db     475 FVCVINGEEHEWTHRAIFRFVYRHEDELELEADDPLLVELQDEDYWYEAYNERTGARGVF 534

Qy     540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
Db     535 PFYYAIEVTKEGEHMAALAKN---------------------------------------- 555

Qy     600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
Db     556 --------HCVLEISVRGIKIGVKADDAKEAKGNKCSHLFQLKNISFCMYHPKNNKYFNF 607

Qy     660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
Db     608 ITKHPADHQFACHVFVSERSTKALAESVSRAFQQFYKQTVEYTCPTEDVYLE 659


RESULT 14
ABB04790
ID   ABB04790 standard; Protein; 659 AA.
XX
AC   ABB04790;
XX
DT   13-MAR-2002  (first entry)
XX
DE   LDL receptor binding protein JIP-1 SEQ ID NO:13.
XX
KW   Low density lipoprotein receptor binding protein; signal transduction;
KW   LDL receptor binding protein; LDL receptor signalling pathway.
XX
OS   Synthetic.
XX
PN   WO200184159-A2.
XX
PD   08-NOV-2001.
XX
PF   24-APR-2001; 2001WO-US13214.
XX
PR   01-MAY-2000; 2000US-0562737.
XX
PA   (TEXA ) UNIV TEXAS SYSTEM.
XX
PI   Herz J,  Gotthardt M;
XX
DR   WPI; 2002-082855/11.
XX
PT   Detecting stress that alters interaction of LDL receptor binding
PT   polypeptide with LDL receptor interaction domain, comprises detecting
PT   difference in stress-biased and unbiased interaction of peptide and
PT   domain in a system  -
XX
PS   Disclosure; Page 38-39; 200pp; English.
XX
CC   The present invention describes a method for detecting a stress that
CC   alters a functional interaction of a low density lipoprotein (LDL)
CC   receptor binding protein (I) with an LDL receptor interaction domain
CC   (II). The method involves introducing a predetermined stress into a
CC   system which provides a stress-biased physical interaction of (I) with
CC   (II), where in the absence of the stress, the system provides an
CC   unbiased interaction of (I) and (II), and detecting the stress-biased
CC   interaction of (I) and (II), where a difference between BI and UI
CC   indicates that the stress alters the interaction of (I) and (II). (I)
CC   is selected from SEMCAP-1, JIP-1, PSD-95, JIP-2, Talin, OMP25, CAPON,
CC   PIP4,5 Kinase, Na channel brain 3, Mint1, ICAP-1 and APC subunit 10.
CC   The method is useful for detecting a stress that alters functional
CC   interaction of LDL receptor binding polypeptide with LDL receptor
CC   interaction domain. The method is useful for detecting and modulating
CC   signal transduction through LDL receptors. ABB04778 to ABB04909
CC   represent LDL receptor binding proteins which are used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   659 AA;

  Query Match             75.1%;  Score 2846;  DB 23;  Length 659;
  Best Local Similarity   78.4%;  Pred. No. 5e-214;
  Matches  558;  Conservative   20;  Mismatches   80;  Indels   54;  Gaps    5;

Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
Db       1 MAARES-GLGGGADSPPAASPFLELHIASPPNFFLTHDISLEEGEDEDLSEITHECGISL 59

Qy      61 QCKDTLSLRPPRAGLLSAGGGGA-GSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
Db      60 QCKITLSLRPPRAKLLSAGSSGSLGSRLQAEMLMMDLIDAAGDNPGAEDDEEEQDDELAA 119

Qy     120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
Db     120 QRPRVGPPKAESNSDPAPR----SQGQTPGTGSGDTYVPKRPTTLNLWPQVPRSQDTYNN 175

Qy     180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
Db     176 NSLGKKHAWQDRVSRSSDPLKTGEQTPEHEHICLSDEFPPQGSPVPTGDRGTSTDSPHRR 235

Qy     240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
Db     236 SAATQMAIPSGPPATAPKGRGHSHRDRLHYQADVRLEMTEEIYLTPVNRPPDPAEPTQTF 295
```

```
Qy    300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
          :|||||| |||||||||:|  |||||| |||||:||| |||||||||| |||||||||  ||
Db    296 MPPTESRRSVSSDPDPASYSVTAGRPHTSISEEDEGFVCLSSPERAEWPGGGWRGSLYEP 355

Qy    360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
          ||||||| |||||||||||||||||| |||||||| |||||||| |||||||| |
Db    356 PPPPRASASSDTSALSYDSVKYTLVVDFHAQLELVSLGPCFGDYSDEHDSATVYDNCISA 415

Qy    420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
          ||||||| |||||||: :|| ||||| ||||||| || ||||||||||:||||||||| |
Db    416 SSPYESAKGEEYEEAPQLRPPTCLSEDMTPDEPDV-FSNKFLNVFMSGQSRSSSAESFRL 474

Qy    480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
          |||:|||| |||||||| |||||||| |||||||| |||||||| |||||||| |||||||
Db    475 FSCVINGESHEQTHRAIFTFVPRHEDELVLEVDDPLLVWLQAEDYWYEYYNMRTGARGAF 534

Qy    540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
          |||||||| |||||||| ||
Db    535 PAYYAIEVDKEPEHMAALEKN---------------------------------------- 555

Qy    600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
                |||||| |||||||| |: |||||| |||||||| |||||||| ||||
Db    556 --------SCVLEISFRGVKIGVKAGDALEAKGNKHSHFFQLKNIIFCGYHPKNNLYFGF 607

Qy    660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          ||||| |||||||| |||||||||:||||||| |||||||||:||||||
Db    608 ITKHPMDHRFACHVFNSEDSTKALAQSVGRAFQQFRKQFVEYTCPSEDIYLE 659
```

RESULT 15
ABB04794
ID ABB04794 standard; Protein; 659 AA.
XX
AC ABB04794;
XX
DT 13-MAR-2002 (first entry)
XX
DE LDL receptor binding protein JIP-1 SEQ ID NO:17.
XX
KW Low density lipoprotein receptor binding protein; signal transduction;
KW LDL receptor binding protein; LDL receptor signalling pathway.
XX
OS Synthetic.
XX
PN WO200184159-A2.
XX
PD 08-NOV-2001.
XX
PF 24-APR-2001; 2001WO-US13214.
XX
PR 01-MAY-2000; 2000US-0562737.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Herz J, Gotthardt M;
XX
DR WPI; 2002-082855/11.
XX
PT Detecting stress that alters interaction of LDL receptor binding
PT polypeptide with LDL receptor interaction domain, comprises detecting
PT difference in stress-biased and unbiased interaction of peptide and
PT domain in a system -
XX
PS Disclosure; Page 44-46; 200pp; English.
XX
CC The present invention describes a method for detecting a stress that
CC alters a functional interaction of a low density lipoprotein (LDL)
CC receptor binding protein (I) with an LDL receptor interaction domain
CC (II). The method involves introducing a predetermined stress into a
CC system which provides a stress-biased physical interaction of (I) with
CC (II), where in the absence of the stress, the system provides an
CC unbiased interaction of (I) and (II), and detecting the stress-biased
CC interaction of (I) and (II), where a difference between BI and UI
CC indicates that the stress alters the interaction of (I) and (II). (I)
CC is selected from SEMCAP-1, JIP-1, PSD-95, JIP-2, Talin, OMP25, CAPON,
CC PIP4,5 Kinase, Na channel brain 3, Mint1, ICAP-1 and APC subunit 10.
CC The method is useful for detecting a stress that alters functional
CC interaction of LDL receptor binding polypeptide with LDL receptor
CC interaction domain. The method is useful for detecting and modulating
CC signal transduction through LDL receptors. ABB04778 to ABB04909
CC represent LDL receptor binding proteins which are used in the
CC exemplification of the present invention.
XX
SQ Sequence 659 AA;

Query Match 75.0%; Score 2845; DB 23; Length 659;
Best Local Similarity 78.6%; Pred. No. 5.9e-214;
Matches 561; Conservative 19; Mismatches 76; Indels 58; Gaps 6;

```
Qy      1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
          |||||| |||||||| |||||||| |||||||| |||||||| |||||||| ||
Db      1 MAERES-ALGGGAASPPDASPFLGLHIESPPNFRLTHFISLEEFEDEGLSEITDECGHSL 59

Qy     61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
          |||||||:|||||||| |  | ||||:|||||||| || ||||||:||||::|:||||
Db     60 QCKDTLSIRPPRAGLLSKGSSGSAGSRMQAEMLQMDLNDAAGDTPGAQDDEEEEDDERAA 119

Qy    120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPG--SGDTYRPKRPTTLNLFPQVPRSQDTL 177
          :||| || |||| |:||:| ||      |||| |||||||| |||||||| ||||||
Db    120 QRPGVGPSKAESNQDPATRSQG------QGPGTVSGDTYRPKRWTTLNLFPQVYRSQDTL 173

Qy    178 NNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPC 237
          |||:|||||||| |||||||| |||||||| |||:|||||  | || ||| ||||||
Db    174 NNNALGKKHSWQDDVSRSSSPLKEGEQTPPHEHFCLSDELPPQHSPVPTQDRGISTDSPC 233

Qy    238 RRSTATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTS 297
          ||| ||||||| | ||  ||||| |||||||| |||||||| |||||||| |||||
Db    234 RRSKATQMAPPSGLPATAPGGRGMSHRDRIHYQNDVRLEATEEQYLTPVQRPPRPAEPTS 293

Qy    298 AFLPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLG 357
          |: |||||||||:|||||||   |||||||| |:||||||| |||||||| ||||||
Db    294 TFMSPTESRMSVSTDPDPAAYSVVAGRPHPSISWEDEGFDCLSYPERAEPPGGAWRGSLG 353

Qy    358 EPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCA 417
          ||| ||||||||| |||||||||:|||||||| |||||||| |||||||| ||||||
Db    354 EPPDPPRASLSSDESALSYDSVKFTLVVDEHAQGELVSLRPCFHDYSDESDSAIVYDNCA 413

Qy    418 SVSSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESF 477
          | | ||||||||| ||||:|:|| ||||||||| ||| |||||| |||||||| |||||
Db    414 SASKPYESAIGEELEEAPQPRPPMCLSEDSTPDNPDV-FSKKFLQVFMSGRSRSRSAESF 472

Qy    478 GLFSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARG 537
          |||| :||||| || ||||||||| |||||||| |||||||| |||||||| |||||
Db    473 GLFSSVINGEEHEQVHRAIFRFVPWHEDELELEVYDPLLVELQAADYWYEAYNMDTGARG 532

Qy    538 VFPAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTR 597
          |||| |||||||| |||||||||
Db    533 VFPAEYAIEVTKEPFHMAALAKNSG---------------------------------- 557

Qy    598 RLTVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYF 657
                      |||||||| |||||||:  |||||||| |||||||| ||||||||
Db    558 ------------VLEISVRGVHIGVKADDALIAKGNKCSHFKQLKNISFCGLHPKNNKYF 605

Qy    658 GFITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          | |||||||| |||||||| |||||||| |||||||| |||||||| ||
Db    606 GMITKHPADHRNACHVFVSEDQTKALAESVGSAFQQFYKQFTEYTCPTEDIVLE 659
```

Search completed: December 24, 2002, 08:50:20
Job time : 43 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: December 24, 2002, 08:48:50 ; Search time 26 Seconds
(without alignments)
2628.907 Million cell updates/sec

Title: US-09-966-561-2
Perfect score: 3792
Sequence: 1 MAERESGGLGGGAASPPAAS.....QQFYKQFVEYTCPTEDIYLE 711

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3263.5	86.1	660	2	T03038	probable inhibitor
2	246.5	6.5	315	2	A88043	protein C13A10.3 [
3	195.5	5.2	705	2	A35363	synapsin I splice
4	194.5	5.1	2142	2	B35098	MHC class III hist
5	191	5.0	1872	2	S36152	MHC class III hist
6	190.5	5.0	1870	2	S37671	MHC class III hist
7	184.5	4.9	706	2	E30411	synapsin Ia - bovi
8	183	4.8	1733	2	S27939	tensin - chicken
9	182	4.8	1634	2	T26517	hypothetical prote
10	179.5	4.7	4957	2	T03455	ALR protein - huma
11	179.5	4.7	5262	2	T03454	ALR protein - huma
12	179	4.7	1744	2	A54970	tensin, cardiac mu
13	179	4.7	1792	2	A57075	tensin - chicken (
14	177.5	4.7	1240	2	JC5209	insulin receptor s
15	177.5	4.7	1242	2	JS0670	insulin receptor s
16	174.5	4.6	1077	2	A44067	serine-rich protei
17	174	4.6	1184	2	G01763	atrophin-1 - human
18	173.5	4.6	403	2	S52796	prpL2 protein - hu
19	173.5	4.6	1110	2	T19673	hypothetical prote
20	171.5	4.5	1279	2	T18312	hypothetical prote
21	170.5	4.5	440	2	JC7807	Wiskott-Aldrich sy
22	170.5	4.5	1664	2	T18262	S-layer protein -
23	170	4.5	3149	1	QQBE8	BPLF1 protein - hu
24	169.5	4.5	1184	2	S50832	atrophin-1 - human
25	169	4.5	1231	2	S30185	insulin receptor s
26	167.5	4.4	1958	2	B40505	hypothetical prote
27	167	4.4	1262	2	T13353	protein stn-B - fr
28	166.5	4.4	1235	1	S16948	insulin receptor s
29	166	4.4	704	2	A30411	synapsin Ia - rat
30	165.5	4.4	548	2	S59133	ETS2 repressor fac
31	165	4.4	1291	2	T00019	period protein hom
32	164.5	4.3	379	2	S31719	proline-rich prote
33	164	4.3	580	2	T43481	probable mucin DKF
34	163.5	4.3	649	2	T46500	hypothetical prote
35	163.5	4.3	780	2	T00366	hypothetical prote
36	162.5	4.3	295	2	B48013	proline-rich prote
37	162.5	4.3	476	2	T27051	hypothetical prote
38	162	4.3	990	2	I51618	nucleolar phosphop
39	162	4.3	2649	2	T51023	hypothetical prote
40	161	4.2	1320	2	JC5630	TCOF1 protein - mo
41	160.5	4.2	1768	2	T13349	parallel sister ch
42	160	4.2	1736	2	T00391	hypothetical prote
43	159.5	4.2	707	2	S60588	drebrin A - rat
44	159.5	4.2	929	2	C96623	hypothetical prote
45	159.5	4.2	1110	2	T29327	hypothetical prote

ALIGNMENTS

RESULT 1
T03038
probable inhibitor protein JIP-1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
C;Accession: T03038
R;Dickens, M.; Rogers, J.S.; Cavanagh, J.; Raitano, A.; Xia, Z.; Halpern, J.R.; Green
Science 277, 693-696, 1997
A;Title: A cytoplasmic inhibitor of the JNK signal transduction pathway.
A;Reference number: Z14833; MUID:97382313; PMID:9235893
A;Accession: T03038
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-660 <DIC>
A;Cross-references: EMBL:AF003115; NID:g2316073; PIDN:AAB66317.1; PID:g2316074
C;Genetics:
A;Note: JIP-1
C;Function:
A;Description: binds to the JNK protein kinase and inhibits JNK signal transduction p

Query Match 86.1%; Score 3263.5; DB 2; Length 660;
Best Local Similarity 87.2%; Pred. No. 2.6e-175;
Matches 621; Conservative 12; Mismatches 26; Indels 53; Gaps 4;

```
Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAERES-GLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 59

Qy      61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
           |||||||||||||||||  | ||||||||||||||| ||||||||||::|:| ||
Db      60 QCKDTLSLRPPRAGLLSAGSSGSAGSRLQAEMLQMDLIDAAGDTPGAEDDEEEEDDELAA 119

Qy     120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
           :||| ||||||| |:|| |    |||| | ||||||||||||||||||||||||||||||
Db     120 QRPGVGPPKAESNQDPAPR----SQGQGPGTGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 175

Qy     180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
           |||||||||||||||||||||||||||||||:|||||  | || ||||||||||||
Db     176 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSDELPPQGSPVPTQDRGTSTDSPCRR 235

Qy     240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
           | ||||||| |||| ||||||||||||||||||||||||||||||||||||| ||||| |
Db     236 SAATQMAPPSGPPATAPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDPAEPTSTF 295

Qy     300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
           :|||||||||||||||  ||||||||||:|||||||||||||||||||||||||||||||
Db     296 MPPTESRMSVSSDPDPAAYSVTAGRPHPSISEEDEGFDCLSSPERAEPPGGGWRGSLGEP 355

Qy     360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     356 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASA 415
```

```
Qy    420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
          ||||||||||||||||:|:|| ||||||||||||||||||||||||||||||||||||||
Db    416 SSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 475

Qy    480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
          |||:||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    476 FSCVINGEEHEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 535

Qy    540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
          ||||||||||||||||||||
Db    536 PAYYAIEVTKEPEHMAALAKN---------------------------------------- 556

Qy    600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
                  ||||||||||||||||||: ||||||||||||||||||||||||||||||||||
Db    557 --------SCVLEISVRGVKIGVKADDALEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 608

Qy    660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 660


RESULT 2
A88043
protein C13A10.3 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: A88043
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: A88043
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-315 <STO>
A;Cross-references: GB:chr_II; PIDN:AB37940.1; PID:g1707200; GSPDB:GN00020; CESP:C13A10.
C;Genetics:
A;Gene: C13A10.3
A;Map position: 2

  Query Match             6.5%;  Score 246.5;  DB 2;  Length 315;
  Best Local Similarity   36.8%;  Pred. No. 6.7e-07;
  Matches   56;  Conservative   27;  Mismatches   62;  Indels    7;  Gaps    3;

Qy    475 ESFGLFSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTG 534
          :| |: ||    : |  |||    | ||| ||| ||: | : |:  |:|:|    |:|||
Db     31 DSSGVSSC-TTSDSQNPTHRVQSAFHPRHPDELLLEIGDAVHVDRTADDHWSYGTNLRTG 89

Qy    535 ARGVFPAYYAIEVTKEPE-HMAALAKNSDWV-----DQFRVKFLGSVQVPYHKGNDVLCA 588
            |:|||    |:    |  : ||  |:  :     | | :  | |::| :|||||||
Db     90 QSGIFPASIVCEIDLVEEICLGALPTNATKILSGDRDTFYLTMLASIEVAHHKGNDVLTQ 149

Qy    589 AMQKIATTRRLTVHFNPPSSCVLEISVRGVKI 620
          || |: :  : :     | : ::||| ||: :
Db    150 AMNKVLSMYKNSEEIIVPQTVLMEISFRGIHV 181


RESULT 3
A35363
synapsin I splice form a - human
C;Species: Homo sapiens (man)
C;Date: 17-Aug-1990 #sequence_revision 17-Aug-1990 #text_change 01-Dec-2000
C;Accession: A35363; B35363; A35805
R;Suedhof, T.C.
J. Biol. Chem. 265, 7849-7852, 1990
A;Title: The structure of the human synapsin I gene and protein.
A;Reference number: A35363; MUID:90243651; PMID:2110562
A;Accession: A35363
A;Molecule type: DNA
A;Residues: 1-705 <SUE>
A;Cross-references: GB:M58371; GB:J05431
A;Accession: B35363
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-659,'KASPAQAQP' <SU2>
A;Cross-references: GB:M58378; GB:J05431
R;Sauerwald, A.; Hoesche, C.; Oschwald, R.; Kilimann, M.W.
J. Biol. Chem. 265, 14932-14937, 1990
A;Title: The 5'-flanking region of the synapsin I gene. A G+C-rich, TATA- and CAAT-le
A;Reference number: A35805; MUID:90368667; PMID:2118519
A;Accession: A35805
A;Molecule type: DNA
A;Residues: 1-125 <SAU>
A;Cross-references: GB:M55301; NID:g338655; PIDN:AAA60608.1; PID:g553654; GB:J05630
C;Genetics:
A;Gene: GDB:SYN1
A;Cross-references: GDB:119606; OMIM:313440
A;Map position: Xp11.23-Xp11.23
C;Keywords: actin binding; alternative splicing; phosphoprotein

  Query Match             5.2%;  Score 195.5;  DB 2;  Length 705;
  Best Local Similarity   25.8%;  Pred. No. 0.0012;
  Matches   73;  Conservative   25;  Mismatches  128;  Indels   57;  Gaps   10;

Qy    100 TGDTPGAEDDEEDDDEERAARRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPK 159
          |  |     ::    :    :|| || :    :    |||      || :|   |:
Db    448 TSQQPAGPPAQQRPPPQGGPPQPGPGPQRQGPPLQQRPPPQGQQHLSGLGPPAGSPL-PQ 506

Qy    160 RPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPP 219
          |  :   || | ||        |          | | |:  |   ||      |
Db    507 RLPSPTSAPQQPASQAAPPTQGQG----------RQSRPVAGGPGAPPAARPPASPSPQR 556

Qy    220 QSGPAPTTDRGTSTDSPCRRSTATQMAPPG-----GPPAAPPGGRGHSHRDRIHYQADVR 274
          |:|| |   | ||   |    :   ||||     |||  |||  | :
Db    557 QAGP-PQATRQTSVSGPAPPKASG--APPGGQQRQGPPQKPPGPAGPTR----------- 602

Qy    275 LEATEEIYLTPVQR--PPDAAEPT-----SAFLPPTESRMSVSSDPDPAAYPSTAGRPHP 327
            :|::    || |  ||  :|       |  |  :    | |  | |  :  | |||
Db    603 -QASQ---AGPVPRTGPPTTQQPRPSGPGPAGAPKPQLAQKPSQDVPPPATAAAGGPPHP 658

Qy    328 SISEEEEGFDCLSSPERAEPPGGGWRGSLGEPPPPPRASLSSD 370
            ::: :   :  : ||           | |||  ||| |
Db    659 QLNKSQSLTNAFNLPE---------------PAPPRPSLSQD 685


RESULT 4
B35098
MHC class III histocompatibility antigen HLA-B-associated protein 2 [imported] - huma
C;Species: Homo sapiens (man)
C;Date: 10-Aug-1990 #sequence_revision 06-Nov-1992 #text_change 24-Aug-2001
C;Accession: B35098
R;Banerji, J.; Sands, J.; Strominger, J.L.; Spies, T.
Proc. Natl. Acad. Sci. U.S.A. 87, 2374-2378, 1990
A;Title: A gene pair from the human major histocompatibility complex encodes large pr
A;Reference number: A35098; MUID:90192810; PMID:2156268
A;Accession: B35098
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-2142 <BAN>
A;Cross-references: GB:M33509; NID:g179338; PIDN:AAA35585.1; PID:g179339; GB:M31293
A;Note: the authors translated the codon AGT for residue 97 as Gly
C;Superfamily: collagen alpha 1(IV) chain

  Query Match             5.1%;  Score 194.5;  DB 2;  Length 2142;
  Best Local Similarity   23.2%;  Pred. No. 0.0053;
  Matches  118;  Conservative   33;  Mismatches  170;  Indels  187;  Gaps   22;

Qy     13 AASPPAASPFLGLHIASPPNFRLTHDISLEEFED-EDLSEITDECGISLQCKDTLSLRPP 71
          | :|| |:        ||       | : ||:|   : |: | |     :    : | |
Db   1281 APAPPRAA------AKSPDLSNQNSDQANEEWETASESSDFTSERRGDKEAPPPVLLTPK 1334

Qy     72 RAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAARRPGAGPPKAES 131
            |    |||||   : | | : ||     |        :    ||  |||| |     |
Db   1335 AVGTPGGGGGGAVPGISA-MSRGDLSQRAKDLSKRSFSSQRPGMERQNRRPGPGGKAGSS 1393
```

```
Qy    132 GQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDR 191
Db   1394 GS--SSGGGGGGPGGRTGPGRGD----------------------------KRSWPSP 1421

Qy    192 VSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRRSTA--------- 242
Db   1422 KNRSRPP----EERPP------GLPLPP---PPPSSSAVFRLDQVIHSNPAGIQQALAQL 1468

Qy    243 -----TQMAPPG------GPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPD 291
Db   1469 SSRQGSVTAPGGHPRHKPGPPQAPQG----------------------------PSPRPPT 1501

Qy    292 AAEPTSAFLPPTESRMSVSSDP----------------DPAAY----PSTAGRP--HPS 328
Db   1502 RYEPQRV-------NSGLSSDPHFEEPGPMVRGVGGTPRDSAGVSPFPPKRRERPPRKPE 1554

Qy    329 ISEEEEGFDCLSSPERAEPPGGGWRGSLGEPPPPPRASLSSDTSALS---YDSVKYTLVV 385
Db   1555 LLQEE------SLP---PPHSSGFLGSKPEGPGPQAESRDTGTEALTPHIWNRL------ 1599

Qy    386 DEHAQLELVSLRPCFGDYSDESDSATVYDNCASVSSPYESAIGEEYEEAPRPQPPACLSE 445
Db   1600 --HTATSRKSYRP-----------TSMEPWMEPLSPFEDVAGTEMSQS----------- 1634

Qy    446 DSTPDEPDVHFSKKFLNVFMSGRSRSSS 473
Db   1635 DSGVD-------------LSGDSQVSS 1648
```

RESULT 5
S36152
MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - human
C;Species: Homo sapiens (man)
C;Date: 06-Jun-1995 #sequence_revision 17-Nov-1995 #text_change 15-Sep-2000
C;Accession: S36152
R;Iris, F.J.M.; Bougueleret, L.; Prieur, S.; Caterina, D.; Primas, G.; Perrot, V.; Jurka
Nature Genet. 3, 137-145, 1993
A;Title: Dense Alu clustering and a potential new member of the NFkappaB family within a
A;Reference number: S36152; MUID:93272029; PMID:8499947
A;Accession: S36152
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1872 <IRI>
A;Cross-references: EMBL:Z15025
A;Note: in the authors' translation residues 32-34 are shown after residue 4 and, consec
A;Note: the authors translated the codon AAT for residue 1000 as His
C;Genetics:
A;Introns: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1; 65
C;Superfamily: collagen alpha 1(IV) chain

Query Match 5.0%; Score 191; DB 2; Length 1872;
Best Local Similarity 22.7%; Pred. No. 0.0071;
Matches 125; Conservative 34; Mismatches 180; Indels 212; Gaps 24;

```
Qy      3 ERES---GGLGGGAASPPAASP-----FLGLHIA---------SPPNFRLTHDISLEEFE 45
Db   1260 ERENAARGSEGKPSLTLPASAPGPEEALTTVTVAPAPRRAAAKSPDLSNQNSDQANEEWE 1319

Qy     46 D-EDLSEITDECGISLQCKDTLSLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTP 104
Db   1320 TASESSDFTSERRGDKEAPPPVLLTPKAVGTPGGGGGGAVPGISA-MSRGDLSQRAKDLS 1378

Qy    105 GAEDDEEDDDEERAARRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTL 164
Db   1379 KRSFSSQRPGMERQNRRPGPGGKAGSSGS--SSGGAGGGPGGRTGPGRGD---------- 1426

Qy    165 NLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPA 224
Db   1427 -------------------KRSWPSPKNRSRPP----EERPP------GLPLPP---PP 1453

Qy    225 PTTDRGTSTDSPCRRSTA--------------TQMAPPG------GPPAAPPGGRGHSHR 264
```

```
Db   1454 PSSSAVFRLDQVIHSNPAGIQQALAQLSSRQGSVTAPGGHPRHKPGPPQAPQG------- 1506

Qy    265 DRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAFLPPTESRMSVSSDP----------- 313
Db   1507 -------------------PSPRPPTRYEPQRV-------NSGLSSDPHFEEPGPMVRG 1539

Qy    314 ------DPAAY----PSTAGRP--HPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEPPP 361
Db   1540 VGGTPRDSAGVSPFPPKRRERPPRKPELLQEE------SLP---PPHSSGFLGSKPEGPG 1590

Qy    362 PPRASLSSDTSALS---YDSVKYTLVVDEHAQLELVSLRP---------------CFGDY 403
Db   1591 PQAESRDTGTEALTPHIWNRL--------HTATSRKSYRPSSMEPWMEPLSPFEDVAGTE 1642

Qy    404 SDESDSATVYDNCASVSS--------------------------------PYESAIGEE 430
Db   1643 MSQSDSGVDLSGDSQVSSGPCSQRSSPDGGLKGAAEGPPKRPGGSSPLNAVPCEGPPG-- 1700

Qy    431 YEEAPRPQPPA 441
Db   1701 -SEPPRRPPPA 1710
```

RESULT 6
S37671
MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - hu
C;Species: Homo sapiens (man)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 15-Sep-2000
C;Accession: S37671
R;Bougueleret, L.
submitted to the EMBL Data Library, August 1992
A;Reference number: S37671
A;Accession: S37671
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1870 <BOU>
A;Cross-references: EMBL:Z15025; NID:g29374; PID:g29375
C;Genetics:
A;Map position: 6p21.3
A;Introns: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1;
C;Superfamily: collagen alpha 1(IV) chain

Query Match 5.0%; Score 190.5; DB 2; Length 1870;
Best Local Similarity 23.0%; Pred. No. 0.0076;
Matches 127; Conservative 32; Mismatches 179; Indels 213; Gaps 25;

```
Qy      3 ERES---GGLGGGAASPPAASP-----FLGLHIA---------SPPNFRLTHDISLEEFE 45
Db   1259 ERENAARGSEGKPSLTLPASAPGPEEALTTVTVAPAPRRAAAKSPDLSNQNSDQANEEWE 1318

Qy     46 D-EDLSEITDECGISLQCKDTLSLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTP 104
Db   1319 TASESSDFTSERRGDKEAPPPVLLTPKAVGTPGGGGGGAVPGISA-MSRGDLSQRAKDLS 1377

Qy    105 GAEDDEEDDDEERAARRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTL 164
Db   1378 KRSFSSQRPGMERQNRRPGPGGKAGSSGS--SSGGAGGGPGGRTGPGRGD---------- 1425

Qy    165 NLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPP----- 219
Db   1426 -------------------KRSWPSPKNRSRPP----EERPP------GLPLPPPPPSS 1455

Qy    220 ---------QSGPAPTTDRGTSTDSPCRRSTATQMAPPG------GPPAAPPGGRGHSHR 264
Db   1456 SVFRLDQVIHSNPAGIQQALAQLSS--RQGSVT--APGGHPRHKPGPPQAPQG------- 1504

Qy    265 DRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAFLPPTESRMSVSSDP----------- 313
Db   1505 -------------------PSPRPPTRYEPQRV-------NSGLSSDPHFEEPGPMVRG 1537

Qy    314 ------DPAAY----PSTAGRP--HPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEPPP 361
Db   1538 VGGTPRDSAGVSPFPPKRRERPPRKPELLQEE------SLP---PPHSSGFLGSKPEGPG 1588
```

```
Qy     362 PPRASLSSDTSALS---YDSVKYTLVVDEHAQLELVSLRP----------CFGDY 403
Db    1589 PQAESRDTGTEALTPHIWNRL--------HTATSRKSYRPSSMEPWMEPLSPFEDVAGTE 1640

Qy     404 SDESDSATVYDNCASVSS--------------------------------PYESAIGEE 430
Db    1641 MSQSDSGVDLSGDSQVSSGPCSQRSSPDGGLKGAAEGPPKRPGGSSPLNAVPCEGPPG-- 1698

Qy     431 YEEAPRPQPPA 441
Db    1699 -SEPPRRPPPA 1708


RESULT 7
E30411
synapsin Ia - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 01-Dec-2000
C;Accession: E30411; F30411; A35758
R;Suedhof, T.C.; Czernik, A.J.; Kao, H.T.; Takei, K.; Johnston, P.A.; Horiuchi, A.; Kana
Science 245, 1474-1480, 1989
A;Title: Synapsins: mosaics of shared and individual domains in a family of synaptic ves
A;Reference number: A30411; MUID:89388265; PMID:2506642
A;Accession: E30411
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-706 <SUE>
A;Accession: F30411
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-661,'KA',664,'PAQAQP' <SU2>
R;Hall, F.L.; Mitchell, J.P.; Vulliet, P.R.
J. Biol. Chem. 265, 6944-6948, 1990
A;Title: Phosphorylation of synapsin I at a novel site by proline-directed protein kinas
A;Reference number: A35758; MUID:90216728; PMID:2108963
A;Accession: A35758
A;Molecule type: protein
A;Residues: 532-556 <HAL>
C;Keywords: actin binding; alternative splicing; phosphoprotein
F;551/Binding site: phosphate (Ser) (covalent) (by proline-directed kinase) #status expe

  Query Match             4.9%;  Score 184.5;  DB 2;  Length 706;
  Best Local Similarity   27.3%;  Pred. No. 0.0052;
  Matches   74;  Conservative   23;  Mismatches  105;  Indels   69;  Gaps   12;

Qy     122 PGAGPPKAESG---------QEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPR 172
Db     463 PQGGPPQPGPGPQRQGPPLQQRPTPQGQQHLSGLGPPAGSPLPQRLPSPTSV---PQQPA 519

Qy     173 SQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTS 232
Db     520 SQATPMTQGQG----------RQSRPVAGGPGAPPATRPPASPSPQRQAGP-PQATRQTS 568

Qy     233 T--DSPCRRSTATQMAPPG-----GPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTP 285
Db     569 VSGQAPPKASG----VPPGGQQRQGPPQKPPGPAGPTRQ------------ASQAGPMP 611

Qy     286 VQRPPDAAEP-TSAFLP---PTESRMS--VSSDPDPAAYPSTAGRPHPSISEEEEGFDCL 339
Db     612 RTGPPTTQQPRPSGPGPAGRPTKPQLAQKPSQDVPPPATAAAGGPPHPQLNKSQSLTNAF 671

Qy     340 SSPERAEPPGGGWRGSLGEPPPPPRASLSSD 370
Db     672 NLPE---------------PAPPRPSLSQD 686


RESULT 8
S27939
tensin - chicken
C;Species: Gallus gallus (chicken)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 12-Feb-1999
C;Accession: S27939; S28973
R;Chen, L.B.
submitted to the EMBL Data Library, August 1991
A;Reference number: S27939
A;Accession: S27939
A;Molecule type: mRNA
A;Residues: 1-1733 <CHE>
A;Cross-references: EMBL:M74165; NID:g212751; PID:g212752
R;Weigt, C.; Gaertner, A.; Wegner, A.; Korte, H.; Meyer, H.E.
J. Mol. Biol. 227, 593-595, 1992
A;Title: Occurrence of an actin-inserting domain in tensin.
A;Reference number: S28973; MUID:93021103; PMID:1404377
A;Accession: S28973
A;Molecule type: protein
A;Residues: 862-871,'X',873-875,'A',877-1212 <WEI>
C;Superfamily: SH2 homology
F;1461-1570/Domain: SH2 homology <SH2>

  Query Match             4.8%;  Score 183;  DB 2;  Length 1733;
  Best Local Similarity   20.0%;  Pred. No. 0.018;
  Matches  176;  Conservative   97;  Mismatches  313;  Indels  294;  Gaps   39;

Qy      13 AASPPAASPFLGLHIASPPN------FRLTHDISLEEFEDEDLSEITDECGISLQCKDTL 66
Db     929 ASYSPAGSQ--QLLVSSPPSPTAPAQSQLPHK-GLESYED----------LSRSGEEPL 974

Qy      67 SLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAARRP---- 122
Db     975 NLE----GLVAHRVAGVQSREKSP--EESTVPARRRTPSDSHYEKSSPEPGSPRSPTVLS 1028

Qy     123 ----------GAGPPK---------AESGQEPASRGQGQSQGQSQGPGSGDT------YR 157
Db    1029 PEVVSTIAANPGGRPKEPHLHSYKEAFEEMESASPSSLTSGGVRSPPGLAKTPLSALGLK 1088

Qy     158 PKRPTTLNLFP----------QVPRSQDT-----LNNNSLGKKHSWQDR 191
Db    1089 PHNPADILLHPVGEPRSYVESVARTATTGRAGNLPAAQPVGLEVPARNGAFGNSFTVPSP 1148

Qy     192 VSRSSSPLKT----------GEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDS 235
Db    1149 VS-TSSPIHSVDGASLRSYPSEGSPHGTVTPPHA-VAETAYRSPMVSQTPSAHSSYQTSS 1206

Qy     236 P----------CRRSTATQM 245
Db    1207 PSSFQAGTLGSPYASPDYPDGRGGFQPDPQARQQPQVSVVGVHALPGSPRTLHRTVATNT 1266

Qy     246 AP----------PGGP--------PAAPPGGRGHSHRDRIHYQADVRLEATEEI 281
Db    1267 PPSPGFGRRAANPAVASVPGSPGLGRHTVSPHAPPGSPSLAR----HQMAAVPPGSPMYG 1322

Qy     282 YLTPVQRPPDAAEPTSA--FLPPTESRMSV----------SSDPDPAAYPSTAGRPHP 327
Db    1323 YSSPEERRPTLSRQSSASGYQPPSTPSFPVSPAYYPGTSTPHSSSPDSAAYRQGSPTPQP 1382

Qy     328 SISEEEEGFDCLSSPERAEPPGGGWRGSL-------GEPPPPPRASLSSDTSALSYDSVK 380
Db    1383 ALPEKRR----MSAGERS--------NSLPNYATVNGKASSPLSSGMSSPSSG-SAVAFS 1429

Qy     381 YTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASVSSP----YESAIGEEYEEAPR 436
Db    1430 HTL----------PDFSKFSMPDISPETRANVKFVQDTSKYWYKPDISREQAIA-- 1473

Qy     437 PQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLFSCIINGEEQEQTHRAI 496
Db    1474 ------LLKDREPGAFIIRDSHSFRGAYGLAMKVASPPP-----TVMQQNKKGDITNELV 1522

Qy     497 FRFV----PRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFPAYYAIEVTKEPE 552
Db    1523 RHFLIETSPRGV-KLKGCPNEPNFGCLSALVYQHSIMPLALPCKLVIPDRDPMEEKKD-- 1579

Qy     553 HMAALAKNS--DWVDQ---FRVKFLGSVQVPYHKGNDVLCAAMQKI----ATTRRLTVHF 603
Db    1580 --AASTTNSATDLLKQGAACNVLFINSVEMESLTGPQAISKAVAETLVADPTPTATIVHF 1637
```

```
Qy    604 NPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQ----LKNISFCGYHPKNNKY--- 656
                ::| :|: :            |:   ||:    |  ::||   |:  |:
Db   1638 --------KVSAQGITL----------TDNQRKLFFRRHYPLNTVTFCDLDPQERKWTKT 1679

Qy    657 --------FGFIT-KHPADHRFACHVFVSEDSTKALAESV 687
                  |||:  |  :     ||:|   |  :  |  |
Db   1680 DGSGPAKLFGFVARKQGSTTDNVCHLFAELDPDQPAAAIV 1719


RESULT 9
T26517
hypothetical protein Y18D10A.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26517
R;Harris, B.
submitted to the EMBL Data Library, December 1998
A;Reference number: Z20226
A;Accession: T26517
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1634 <WIL>
A;Cross-references: EMBL:AL034393; PIDN:CAA22308.1; CESP:Y18D10A.1
A;Experimental source: clone Y18D10A
C;Genetics:
A;Gene: CESP:Y18D10A.1
A;Introns: 7/3; 44/3; 106/3; 193/1; 500/3; 533/3; 560/3; 590/3; 709/3; 1218/3; 1318/2; 1

  Query Match             4.8%;  Score 182;  DB 2;  Length 1634;
  Best Local Similarity   21.5%;  Pred. No. 0.019;
  Matches   98;  Conservative   57;  Mismatches  168;  Indels  132;  Gaps   18;

Qy     72 RAGLLSAGGGGAGS----------------RLQAEMLQMDLIDATGDTPGAEDDEEDDD 114
          :  || | |    |                  || ||:: : :
Db    378 KKALLGAQGAAGASEPGSSSSIHGKKGKTVRELQTEMMKGESV---------------- 420

Qy    115 EERAARRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQ 174
          :|:|||       :||     || |: |: | :  | : :   |  |    |    | |
Db    421 KEKAARM------RAE-----ASAGRSQAPGPAPAPAASELQDP--PQDFGLSMSDPGSD 467

Qy    175 DTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTD 234
              :   :||     |:| |: |     ||  :    | ::||   |:|:|
Db    468 SETEEHQKQEKHI-PAMVTRRSARLSALPVTP--KKASSSSKMPPPPSPSPST------- 517

Qy    235 SPCRRS----TATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPP 290
           | ||     | : |:      |  |  ||   | |    | |  | ||     |:   |
Db    518 -PGRRGRRPRTLSTMSMEPAAAAVTPAPRG---RPRSRSAAKVS-ENTE-----PLSEAP 567

Qy    291 DAAEPTSAFLPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCL---SSPERAEP 347
           |        | : | ||:: | :|:   : | |   : |:|||  |      |||: :
Db    568 SAPVKRGRGRPRSRSTMSITEDSEPSTSSTAAKRSKRAESDEEEEQDLKLTNKSPEKPKK 627

Qy    348 PGGGWRGSLGE------------------PPPP----------------------PRAS 366
          |      ::|:                    | ||                        |: :
Db    628 PSKTTEETVGDVLKKRLRDTAKTTATVIHTPGPPLRTRKMERMRAPTAVTSSKKEKPKNA 687

Qy    367 LSSDTSALSYDSVKYTLVVDEHAQLELVSLR-------PCFGDYSDESDSATVYDNCASV 419
           |:|:|    :    |::::|   |:|           |  :  ||   |
Db    688 GSADSSINEEEHEDETMILEEQT-LDLPQQTSQQEPRISCGSELLDEQFDA--------- 737

Qy    420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDV 454
          |  :   :    |    | ||   : ::: : | :
Db    738 SEEHSGTVPSAPELTKNPAPPVPEASEASAEPPKI 772


RESULT 10
T03455
ALR protein - human
C;Species: Homo sapiens (man)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C;Accession: T03455
R;Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
Oncogene 15, 549-560, 1997
A;Title: Structure and expression pattern of human ALR, a novel gene with strong homo
A;Reference number: Z14954; MUID:97388474; PMID:9247308
A;Accession: T03455
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-4957 <PRA>
A;Cross-references: EMBL:AF010404; NID:g2358286; PIDN:AAC51735.1; PID:g2358287
C;Genetics:
A;Gene: ALR
A;Map position: 12
C;Superfamily: human ALR protein
C;Keywords: alternative splicing

  Query Match             4.7%;  Score 179.5;  DB 2;  Length 4957;
  Best Local Similarity   21.8%;  Pred. No. 0.099;
  Matches  110;  Conservative   40;  Mismatches  166;  Indels  189;  Gaps   24;

Qy      1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEI-TDECGIS 59
          :|  || |:||| || |         : ||| |           |     |:  :| | |
Db   1675 LAVPESPGVGGGKASEP---------LLSPPPFG----------ESRKALEVKKEELGAS 1715

Qy     60 LQCKDTLSLRPPRAGLLSAGGGG---AGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEE 116
             : |  ||  | : :   |    |  |:              ||         |
Db   1716 -----SPSYGPPNLGFVDSPSSGTHLGGLELK-------------TP---------DVF 1747

Qy    117 RAARRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLN-LFPQVPRSQD 175
          :|   | |                || : | || |   ||: |     | |  |   |
Db   1748 KAPLTPRA-----------------SQVEPQSPGLG--LRPQEPPPAQALAPSPPSHPD 1787

Qy    176 TLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGP----------- 223
            |    :        :  ||    | || |  |    |||:| |
Db   1788 IFRPGSYTDPY--------AQPPLTPRPQPPPPESCC---ALPPRSLPSDPFSRVPVSPQ 1836

Qy    224 ------APTTDRGTSTDSPCRRSTATQMAPP---GGPPAAPPGGRGHSHRDRIHYQADVR 274
                :| | |  | :: |      :   |     ||: |        ||
Db   1837 SQSSSQSPLTPRPLSAEAFCPSPVTPRFQSPDPYSRPPSRP------QSRDP-------- 1882

Qy    275 LEATEEIYLTPVQRPPDAAEPTSAFLPPTESRMSVSSDPDPAAYPS-TAGRPHPSISEEE 333
                |: :||    |  ||   : :  |: :   ||| |:  ||  |  :   :
Db   1883 --------FAPLHKPPRPQPPEVAFKAGSLAHTSLGAGGFPAALPAGPAGELHAKVPSGQ 1934

Qy    334 EGFDCLSSPERAEPPGGG-----------WRGSLGEP---PPPPRASLSSDTSALSY-- 376
                |    || |             :  ::|||   || |:  |       |:
Db   1935 -------PPNFVRSPGTGAFVGTPSPMRFTFPQAVGEPSLKPPVPQPGLPPPHGINSHFG 1987

Qy    377 --------DSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASVSSPYESAIG 428
                 |  ||:                          :    |   |   : |
Db   1988 PGPTLGKPQSTNYTVATGN------------------------FHPSGSPLGPSSGSTG 2022

Qy    429 EEYEEAPRPQPPACLSEDSTPDEPD 453
          | |  :|  :||: |     |  ||
Db   2023 ESYGLSPL-RPPSVL----PPPAPD 2042


RESULT 11
T03454
ALR protein - human
C;Species: Homo sapiens (man)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C;Accession: T03454
R;Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Ya
Oncogene 15, 549-560, 1997
A;Title: Structure and expression pattern of human ALR, a novel gene with strong homo
A;Reference number: Z14954; MUID:97388474; PMID:9247308
A;Accession: T03454
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-5262 <PRA>
A;Cross-references: EMBL:AF010403; NID:g2358284; PIDN:AAC51734.1; PID:g2358285
C;Genetics:
A;Gene: ALR
A;Map position: 12
C;Superfamily: human ALR protein
```

```
C;Keywords: alternative splicing

  Query Match             4.7%;  Score 179.5;  DB 2;  Length 5262;
  Best Local Similarity  21.8%;  Pred. No. 0.11;
  Matches  110;  Conservative   40;  Mismatches  166;  Indels  189;  Gaps   24;

Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEI-TDECGIS 59
           :|  || |:||| || |         : ||| |           |     |:  :| | |
Db    1980 LAVPESPGVGGGKASEP---------LLSPPPFG----------ESRKALEVKKEELGAS 2020

Qy      60 LQCKDTLSLRPPRAGLLSAGGGG---AGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEE 116
                : |  ||  | : :   |    |  |:               ||         |
Db    2021 -----SPSYGPPNLGFVDSPSSGTHLGGLELK-------------TP---------DVF 2052

Qy     117 RAARRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLN-LFPQVPRSQD 175
           :|   | |                  || : | || |   ||: |     | |  |   |
Db    2053 KAPLTPRA-----------------SQVEPQSPGLG--LRPQEPPPAQALAPSPPSHPD 2092

Qy     176 TLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGP----------- 223
                |    :        :  ||    | || |  |    |||:| |
Db    2093 IFRPGSYTDPY--------AQPPLTPRPQPPPPESCC---ALPPRSLPSDPFSRVPVSPQ 2141

Qy     224 ------APTTDRGTSTDSPCRRSTATQMAPP---GGPPAAPPGGRGHSHRDRIHYQADVR 274
                 :| | |  | :: |      :   |     ||: |         ||
Db    2142 SQSSSQSPLTPRPLSAEAFCPSPVTPRFQSPDPYSRPPSRP------QSRDP-------- 2187

Qy     275 LEATEEIYLTPVQRPPDAAEPTSAFLPPTESRMSVSSDPDPAAYPS-TAGRPHPSISEEE 333
                   |: :||    |  ||   : :  |: :   ||| |:  ||  |  :   :
Db    2188 --------FAPLHKPPRPQPPEVAFKAGSLAHTSLGAGGFPAALPAGPAGELHAKVPSGQ 2239

Qy     334 EGFDCLSSPERAEPPGGG-----------WRGSLGEP---PPPPRASLSSDTSALSY-- 376
                 |     || |             :  ::|||   || |:  |        |:
Db    2240 -------PPNFVRSPGTGAFVGTPSPMRFTFPQAVGEPSLKPPVPQPGLPPPHGINSHFG 2292

Qy     377 --------DSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASVSSPYESAIG 428
                   |  ||:                              :    |   |   : |
Db    2293 PGPTLGKPQSTNYTVATGN-----------------------FHPSGSPLGPSSGSTG 2327

Qy     429 EEYEEAPRPQPPACLSEDSTPDEPD 453
           | |  :|  :||: |     |  ||
Db    2328 ESYGLSPL-RPPSVL----PPPAPD 2347


RESULT 12
A54970
tensin, cardiac muscle - chicken
C;Species: Gallus gallus (chicken)
C;Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 21-Jul-2000
C;Accession: A54970; S38330; S21544
R;Lo, S.H.; An, Q.; Bao, S.; Wong, W.K.; Liu, Y.; Janmey, P.A.; Hartwig, J.H.; Chen, L.B
J. Biol. Chem. 269, 22310-22319, 1994
A;Title: Molecular cloning of chick cardiac muscle tensin. Full-length cDNA sequence, ex
A;Reference number: A54970; MUID:94350987; PMID:8071358
A;Accession: A54970
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: mRNA
A;Residues: 1-1744 <LOA>
A;Cross-references: GB:M96625
R;van de Werken, R.; Gennari, M.; Tavella, S.; Bet, P.; Molina, F.; Lin, S.; Cancedda, R
Eur. J. Biochem. 217, 781-790, 1993
A;Title: Modulation of tensin and vimentin expression in chick embryo developing cartila
A;Reference number: S38330; MUID:94039118; PMID:8223621
A;Accession: S38330
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1469-1744 <VAN>
A;Cross-references: EMBL:X66286; NID:g63802; PIDN:CAA46992.1; PID:g63803
C;Superfamily: SH2 homology
C;Keywords: cardiac muscle; heart
F;1472-1581/Domain: SH2 homology <SH2>

  Query Match             4.7%;  Score 179;  DB 2;  Length 1744;
  Best Local Similarity  19.8%;  Pred. No. 0.031;

  Matches  177;  Conservative   97;  Mismatches  312;  Indels  306;  Gaps   39;

Qy      13 AASPPAASPFLGLHIASPPN------FRLTHDISLEEFEDEDLSEITDECGISLQCKDTL 66
           |:  || |    | ::|||:       :| |   || :||           :|   :: |
Db     928 ASYSPAGSQ--QLLVSSPPSPTAPAQSQLPHK-GLESYED----------LSRSGEEPL 973

Qy      67 SLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAARRP---- 122
           :|     ||::    |  || ::   :   : |  ||      |:   |  : | |
Db     974 NLE----GLVAHRVAGVQSREKSP--EESTVPARRRTPSDSHYEKSSPEPGSPRSPTVLS 1027

Qy     123 ----------GAGPPK---------AESGQEPASRGQGQSQGQSQGPGSGDT------YR 157
                     | ||         |    | ||     | |    ||   |      :
Db    1028 PEVVSTIAANPGGRPKEPHLHSYKEAFEEMESASPSSLTSGGVRSPPGLAKTPLSALGLK 1087

Qy     158 PKRPTTLNLFP------------QVPRS------------------------QDTLNN 179
           |  |  : | |             : |||                        :   |
Db    1088 PHNPADILLHPVGELEGEAGADSEEEPRSYVESVARTATTGRAGNLPAAQPVGLEVPARN 1147

Qy     180 NSLGKKHSWQDRVSRSSSPLKT----------------GEQTPPHEHICLSEELPPQSGP 223
           : |   :    || :|||: :                 | ||||  :  :    |
Db    1148 GAFGNSFTVPSPVS-TSSPIHSVDGASLRSYPSEGSPHGTVTPPHA-VAETAYRSPMVSQ 1205

Qy     224 APTTDRGTSTDSP---------------------------------------------- 236
           |:       | ||
Db    1206 TPSAHSSYQTSSPSSFQAGTLGSPYASPDYPDGRGGFQPDPQARQQPQVSVVGVHALPGS 1265

Qy     237 ---CRRSTATQMAP----------------PGGP--------PAAPPGGRGHSHRDRIHY 269
              |: ||   |                  || |        | ||||    :     |
Db    1266 PRTLHRTVATNTPPSPGFGRRAANPAVASVPGSPGLGRHTVSPHAPPGSPSLAR----HQ 1321

Qy     270 QADVRLEATEEIYLTPVQRPPDAAEPTSA--FLPPTESRMSV-----------SSDPDP 315
           | |   :    | :| :| |  :  :||  : ||:      |            || ||
Db    1322 MAAVPPGSPMYGYSSPEERRPTLSRQSSASGYQPPSTPSFPVSPAYYPGTSTPHSSSPDS 1381

Qy     316 AAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSL-------GEPPPPPRASLS 368
           |||   :  | |:: |:       :|: ||:          ||       |:   |  : :|
Db    1382 AAYRQGSPTPQPALPEKRR----MSAGERS--------NSLPNYATVNGKASSPLSSGMS 1429

Qy     369 SDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASVSSP----YE 424
           | :|  |  :  :||               | |  :|    |      |  |       |:
Db    1430 SPSSG-SAVAFSHTL--------------PDFSKFSMPDISPETRANVKFVQDTSKYWYK 1474

Qy     425 SAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLFSCII 484
           |  |   |         | :|  |    :  |  |   :       :|        : :
Db    1475 PDISREQAIA--------LLKDREPGAFIIRDSHSFRGAYGLAMKVASPPP-----TVMQ 1521

Qy     485 NGEEQEQTHRAIFRFV----PRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFP 540
           :: : |:  :  |:    ||   :|:   ::|    | |  | :    :    : | |
Db    1522 QNKKGDITNELVRHFLIETSPRGV-KLKGCPNEPNFGCLSALVYQHSIMPLALPCKLVIP 1580

Qy     541 AYYAIEVTKEPEHMAALAKNS--DWVDQ---FRVKFLGSVQVPYHKGNDVLCAAMQKI-- 593
              :|  |:    ||   ||  | : |     | |: ||::    |   :  |: :
Db    1581 DRDPMEEKKD----AASTTNSATDLLKQGAACNVLFINSVEMESLTGPQAISKAVAETLV 1636

Qy     594 --ATTRRLTVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQ----LKNISFC 647
               |     |||         ::| :|: :             |:   ||:    |  ::||
Db    1637 ADPTPTATIVHF--------KVSAQGITL----------TDNQRKLFFRRHYPLNTVTFC 1678

Qy     648 GYHPKNNKY-----------FGFIT-KHPADHRFACHVFVSEDSTKALAESV 687
              |:  |:           |||:  |  :     ||:|   |  :  |  |
Db    1679 DLDPQERKWTKTDGSGPAKLFGFVARKQGSTTDNVCHLFAELDPDQPAAAIV 1730


RESULT 13
A57075
tensin - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 21-Jul-2000
C;Accession: A57075
R;Chuang, J.Z.; Lin, D.C.; Lin, S.
J. Cell Biol. 128, 1095-1109, 1995
A;Title: Molecular cloning, expression, and mapping of the high affinity actin-cappin
A;Reference number: A57075; MUID:95204530; PMID:7896874
```

A;Accession: A57075
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1792 <CHU>
A;Cross-references: GB:L06662; NID:g212754; PIDN:AAA73949.1; PID:g212755
C;Superfamily: SH2 homology
F;1520-1629/Domain: SH2 homology <SH2>

Query Match 4.7%; Score 179; DB 2; Length 1792;
Best Local Similarity 19.3%; Pred. No. 0.032;
Matches 171; Conservative 108; Mismatches 311; Indels 298; Gaps 38;

```
Qy     13 AASPPAASPFLGLHIASPPN------FRLTHDISLEEFEDEDLSEITDECGISLQCKDTL 66
Db    976 ASYSPAGSQ--QLLVSSPPSPTAPAQSQLPHK-GLESYED-----------LSRSGEEPL 1021

Qy     67 SLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAARRP---- 122
Db   1022 NLE----GLVAHRVAGVQSREKSP--EESTVPARRRTPSDSHYEKSSPEPGSPRSPTVLS 1075

Qy    123 ----------GAGPPK---------AESGQEPASRGQGQSQGQSQGPGSGDT------YR 157
Db   1076 PEVVSTIAANPGGRPKEPHLHSYKEAFEEMESASPSSLTSGGVRSPPGLAKTPLSALGLK 1135

Qy    158 PKRPTTLNLFP-------------QVPRS-------------------------QDTLNN 179
Db   1136 PHNPADILLHPVGELEGEAGADSEEEPRSYVESVARTATTGRAGNLPAAQPVGLEVPARN 1195

Qy    180 NSLGKKHSWQDRVSRSSSPLKT----------------GEQTPPHEHICLSEELPPQSGP 223
Db   1196 GAFGNSFTVPSPVS-TSSPIHSVDGASLRSYPSEGSPHGTVTPPHA-VAETAYRSPMVSQ 1253

Qy    224 APTTDRGTSTDSP---------------------------------------------- 236
Db   1254 TPSAHSSYQTSSPSSFQAGTLGSPYASPDYPDGRAGFQPDPQARQQPQVSVVGVHALPGS 1313

Qy    237 ---CRRSTATQMAP----------------PGGP--------PAAPPGGRGHSHRDRIHY 269
Db   1314 PRTLHRTVATNTPPSPGFGRRAANPAVASVPGSPGLGRHTVSPHAPPGSPSLAR----HQ 1369

Qy    270 QADVRLEATEEIYLTPVQRPPDAAEPTSA--FLPPTESRMSV-----------SSDPDP 315
Db   1370 MAAVPPGSPMYGYSSPEERRPTLSRQSSASGYQPPSTPSFPVSPAYYPGTSTPHSSSPDS 1429

Qy    316 AAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSL-------GEPPPPPRASLS 368
Db   1430 AAYRQGSPTPQPALPEKRR----MSAGERS--------NSLPNYATVNGKASSPLSSGMS 1477

Qy    369 SDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASVSSPYESAIG 428
Db   1478 SPSSG---SAVAFSHTLPDFSKFSMPDISP--------------ETRANVKFVQDTSKY 1519

Qy    429 EEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLFSCIINGEE 488
Db   1520 WYKPDISRDQAIALL-KDREPGAFIIRDSHSFRGAYGLAMKVASPPP-----TVMQQNKK 1573

Qy    489 QEQTHRAIFRFV----PRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFPAYYA 544
Db   1574 GDITNELVRHFLIETSPRGV-KLKGCPNEPNFGCLSALVYQHSIMPLALPCKLVIPDRDP 1632

Qy    545 IEVTKEPEHMAALAKNS--DWVDQ---FRVKFLGSVQVPYHKGNDVLCAAMQKI----AT 595
Db   1633 MEEKKD----AASTTNSATDLLKQGAACNVLFINSVEMESLTGPQAISKAVAETLVADPT 1688

Qy    596 TRRLTVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQ----LKNISFCGYHP 651
Db   1689 PTATIVHF--------KVSAQGITL----------TDNQRKLFFRRHYPLNTVTFCDLDP 1730

Qy    652 KNNKY-----------FGFIT-KHPADHRFACHVFVSEDSTKALAESV 687
Db   1731 QERKWTKTDGSGPAKLFGFVARKQGSTTENVCHLFAELDPDQPAAAIV 1778
```

RESULT 14
JC5209
insulin receptor substrate 1 - chicken
C;Species: Gallus gallus (chicken)
C;Date: 20-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 16-Jul-1999
C;Accession: JC5209; PC4305
R;Taouis, M.; Taylor, S.I.; Reitman, M.
Gene 178, 51-55, 1996
A;Title: Cloning of the chicken insulin receptor substrate 1 gene.
A;Reference number: JC5209; MUID:97080546; PMID:8921891
A;Accession: JC5209
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1240 <TAO>
A;Cross-references: GB:U43502; NID:g1685084; PIDN:AAC60050.1; PID:g1685085
A;Accession: PC4305
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-5 <TA2>
C;Comment: This protein acts as a docking protein and mediates multiple interactions
C;Genetics:
A;Gene: isr-1
C;Superfamily: insulin receptor substrate IRS-1; pleckstrin repeat homology
C;Keywords: phosphoprotein
F;11-112/Domain: pleckstrin repeat homology <PLK>
F;463,549,610,630,660,730,940,987,1010/Binding site: phosphate (Tyr) (covalent) #stat

Query Match 4.7%; Score 177.5; DB 2; Length 1240;
Best Local Similarity 22.3%; Pred. No. 0.025;
Matches 129; Conservative 59; Mismatches 219; Indels 171; Gaps 26;

```
Qy      2 AERESGGLGG--------------GAASPP------AASP---------FLGLHIASPPN 32
Db    308 ATSPAGGVGGKPSSFRVRESSDGEGTMSPPDQQDGSPVSPSTNRTHAHRHRGSALLQPP- 366

Qy     33 FRLTHDISL---------EEFEDEDLSE-------ITDECGISLQCKDTLSLRPPRAGLL 76
Db    367 --LNHSRSIPMPASRCSPSATSPVSLSSSSTSGHGSTSDCLFPRRSSASVSGSPSDGGFI 424

Qy     77 SAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAARRPGAGPP--KAESGQE 134
Db    425 SSDEYGSSPCDFRSSFRSVTPDSLGHTPPARGEEE---LSNYICMGGKGPSTLTAPNGHY 481

Qy    135 PASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDT----LNNNSLGKKHSWQD 190
Db    482 ILSRG---ANGHRCTPGTG----------LGTSPALAGDEQTSAADLDNRFRKRTH---- 524

Qy    191 RVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRRSTATQMAPPGG 250
Db    525 --SAGTSPTITHQKTPSQSSVASIEEY-------------------------TEMM---- 553

Qy    251 PPAAPPGG--RGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAFLPPTESRMS 308
Db    554 -PAYPPGGGSGGRAWPGRLSAFVPTRSYPEEGLEMHPLERRAGRTSPDSSTLHTDDGYMP 612

Qy    309 VSSDPDPAAYPSTAGR---------PHPSISEEEEGFDCLSSP-ERAEPPG-------GG 351
Db    613 MS----PGWPPVSSGRKGNGDYMSMSPKSVSTPQQITNSMTDPSERVYPNGYMMMSHSGG 668

Qy    352 WRGSLGEPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQL----------ELVSLRPCFG 401
Db    669 CSPDIGGGPSSSSSSSNAVPSGTSYGKLWTNGVGGHHSHVLPHPKPPVESSGGKLLPCTG 728

Qy    402 DYSD---ESDSATVYDNCASVSSPYESAIGEEYEEAPRPQP-------PACLSEDSTPDE 451
Db    729 DYMDMYPVGDSNT--------SDPSLSYYG---PEDPQHKPVLSYYSLPDAFKHPSGPGE 777

Qy    452 PDVHFSKKFLNV-FMSGR----------SRSSSAESFG 478
Db    778 PEEGARHQHLRLSTSSGRLLYAATADDSSSSTSSDTLG 815
```

RESULT 15

```
JS0670
insulin receptor substrate-1 - human
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1992 #sequence_revision 23-Aug-1996 #text_change 16-Jul-1999
C;Accession: I53160; JS0670; PN0678
R;Araki, E.; Sun, X.J.; Haag, B.L.; Chuang, L.; Zhang, Y.; Yang-Feng, T.L.; White, M.; K
Diabetes 42, 1041-1054, 1993
A;Title: Human skeletal muscle insulin receptor substrate-1. Characterization of the cDN
A;Reference number: I53160; MUID:93292738; PMID:8513971
A;Accession: I53160
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1242 <RES>
A;Cross-references: GB:S62539; NID:g386256; PIDN:AAB27175.1; PID:g386257
R;Nishiyama, M.; Wands, J.R.
Biochem. Biophys. Res. Commun. 183, 280-285, 1992
A;Title: Cloning and increased expression of an insulin receptor substrate-1-like gene i
A;Reference number: JS0670; MUID:92181456; PMID:1311924
A;Accession: JS0670
A;Molecule type: mRNA
A;Residues: 1-134,'G',135-361,'R',363-383,'R',385-1242 <NIS>
A;Cross-references: GB:S85963; NID:g246465; PIDN:AAB21608.1; PID:g246466
A;Experimental source: hepatocellular carinoma cell line FOCUS
R;Smith, L.K.; Bradshaw, M.; Croall, D.E.; Garner, C.W.
Biochem. Biophys. Res. Commun. 196, 767-772, 1993
A;Title: The insulin receptor substrate (IRS-1) is a PEST protein that is susceptible to
A;Reference number: PN0678; MUID:94059102; PMID:8240352
A;Accession: PN0678
A;Molecule type: protein
A;Residues: 1-14;338-353;411-439;538-545,'V',547-567;656-697;724-758;932-943;1028-1056;1
C;Comment: This protein appears to be the major substrate for insulin-stimulated tyrosin
o dock various proteins containing the phosphotyrosine-binding Src-homology domain 2 (SH
C;Comment: This protein contains at least 11 PEST regions, which suggests rapid turnover
C;Genetics:
A;Gene: GDB:IRS1
A;Cross-references: GDB:133974; OMIM:147545
A;Map position: 2q36-2q36
A;Introns: #status absent
C;Superfamily: insulin receptor substrate IRS-1; pleckstrin repeat homology
C;Keywords: ATP; phosphoprotein
F;11-113/Domain: pleckstrin repeat homology <PLK>
F;46,465,551,612,632,662,732,941,989,1012/Binding site: phosphate (Tyr) (covalent) #stat
F;78,527,1100,1223/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #
F;161/Binding site: ATP (Lys) #status predicted
F;189,323,441,624,636,795,920,984,1084,1218/Binding site: phosphate (Ser) (covalent) (by
F;300,351,774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status pre

  Query Match             4.7%;  Score 177.5;  DB 2;  Length 1242;
  Best Local Similarity   22.9%;  Pred. No. 0.025;
  Matches  117;  Conservative   57;  Mismatches  213;  Indels  123;  Gaps   22;

Qy      53 TDECGISLQCKDTLSLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEED 112
           | :|    :   ::|  |   | :|:   |:         :    |: | || |  :||
Db     403 TSDCLFPRRSSASVSGSPSDGGFISSDEYGSSPCDFRSSFRSVTPDSLGHTPPARGEEE- 461

Qy     113 DDEERAARRPGAGPP--KAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQV 170
                  | ||    | :|    |||     |    ||:|        |:  |
Db     462 --LSNYICMGGKGPSTLTAPNGHYILSRG---GNGHRCTPGTG------LGTSPALAGDE 510

Qy     171 PRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRG 230
             |   |:|    : |      |  :||  | ::||    :   ||
Db     511 AASAADLDNRFRKRTH------SAGTSPTITHQKTPSQSSVASIEEY------------- 551

Qy     231 TSTDSPCRRSTATQMAPPGGPPAAPPGGR--GHSHRDRIHYQADVRLEATEEIYLTPVQR 288
                   |:|  |  ||    |||  || |   :      |    | : : |::|
Db     552 -----------TEMM-PAYPPGGGSGGRLPGHRHSAFV----PTRSYPEEGLEMHPLER 594

Qy     289 PPDAAEPTSAFLPPTESRMSVSSDPDPAAYPSTAGR---------PHPSISEEEEGFDCL 339
               | |: |   :  | :|    |  |  :||          |:|  ::  : :
Db     595 RGGHHRPDSSTLHTDDGYMPMS----PGVAPVPSGRKGSGDYMPMSPKSVSAPQQIINPI 650

Qy     340 -SSPERAEP-------PGGGWRGSLGEPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQL 391
            |:| :|       | ||    :|  |    :| ::  |  ||  :    |   |: :

Db     651 RRHPQRVDPNGYMMMSPSGGCSPDIGGGPSSSSSSSNAVPSGTSYGKLWTNGVGGHHSHV 710

Qy     392 ----------ELVSLRPCFGDYSDES---DSATVYDNCASVSSPYESAIGEEYEEAPRPQ 438
                        | || ||| : |    || |         ||| :   |    | |: :
Db     711 LPHPKPPVESSGGKLLPCTGDYMNMSPVGDSNT--------SSPSDCYYG---PEDPQHK 759

Qy     439 P-------PACLSEDSTPDEPDVHFSKKFLNV-FMSGR----------SRSSSAESFGLF 480
           |       |        | ||:     : | :   |||          | |:|::| |
Db     760 PVLSYYSLPRSFKHTQRPGEPEEGARHQHLRLSTSSGRLLYAATADDSSSSTSSDSLGGG 819

Qy     481 SCIINGEEQEQT-----HRAIFRFVPRHED 505
             |   |   | :     |: :   :||  |
Db     820 YC---GARLEPSLPHPHHQVLQPHLPRKVD 846


Search completed: December 24, 2002, 08:51:11
Job time : 47 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: December 24, 2002, 08:48:45 ; Search time 36 Seconds
(without alignments)
819.157 Million cell updates/sec

Title: US-09-966-561-2
Perfect score: 3792
Sequence: 1 MAERESGGLGGGAASPPAAS.....QQFYKQFVEYTCPTEDIYLE 711

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3789	99.9	711	1	JIP1_HUMAN	Q9uqf2 homo sapien
2	3549	93.6	707	1	JIP1_MOUSE	Q9wvi9 mus musculu
3	3544.5	93.5	708	1	JIP1_RAT	Q9r237 r c-jun-ami
4	1292.5	34.1	830	1	JIP2_MOUSE	Q9ere9 mus musculu
5	1280.5	33.8	824	1	JIP2_HUMAN	Q13387 homo sapien
6	537.5	14.2	490	1	SP51_DROME	Q9w0k0 drosophila
7	195.5	5.2	705	1	SYN1_HUMAN	P17600 homo sapien
8	194.5	5.1	2142	1	BAT2_HUMAN	P48634 homo sapien
9	193.5	5.1	503	1	WAIP_HUMAN	O43516 homo sapien
10	184.5	4.9	706	1	SYN1_BOVIN	P17599 bos taurus
11	179	4.7	1744	1	TENS_CHICK	Q04205 gallus gall
12	177.5	4.7	1242	1	IRS1_HUMAN	P35568 homo sapien
13	174.5	4.6	1077	1	HLES_DROME	Q02308 drosophila
14	170.5	4.5	415	1	SYN1_CANFA	O62732 canis famil
15	170.5	4.5	1664	1	SLP1_CLOTM	Q06852 clostridium
16	170	4.5	3149	1	TEGU_EBV	P03186 epstein-bar
17	169.5	4.5	1185	1	DRPL_HUMAN	P54259 homo sapien
18	169	4.5	671	1	VINE_HUMAN	O60504 homo sapien
19	169	4.5	1233	1	IRS1_MOUSE	P35569 mus musculu
20	167.5	4.4	2167	1	SHK1_RAT	Q9wv48 rattus norv
21	166.5	4.4	1235	1	IRS1_RAT	P35570 rattus norv
22	166	4.4	704	1	SYN1_RAT	P09951 rattus norv
23	165.5	4.4	548	1	ERF_HUMAN	P50548 homo sapien
24	165	4.4	1083	1	T2D3_HUMAN	O00268 homo sapien
25	165	4.4	1291	1	PER1_MOUSE	O35973 mus musculu
26	165	4.4	1461	1	IE18_PRVIF	P11675 pseudorabie
27	163.5	4.3	551	1	ERF_MOUSE	P70459 mus musculu
28	163.5	4.3	839	1	APB1_RAT	O35430 rattus norv
29	162	4.3	1183	1	DRPL_RAT	P54258 rattus norv
30	159.5	4.2	707	1	DREB_RAT	Q07266 rattus norv
31	159.5	4.2	1324	1	IRS2_HUMAN	Q9y4h2 homo sapien
32	159	4.2	2424	1	CCAA_RABIT	P27884 oryctolagus
33	158.5	4.2	1446	1	IE18_PRVKA	P33479 pseudorabie
34	157.5	4.2	561	1	3BP2_HUMAN	P78314 homo sapien
35	157.5	4.2	1125	1	S24C_HUMAN	P53992 homo sapien
36	157	4.1	234	1	PRPM_HUMAN	P10161 homo sapien
37	156.5	4.1	837	1	APB1_HUMAN	Q02410 homo sapien
38	156.5	4.1	1790	1	SEPA_EMENI	P78621 emericella
39	156	4.1	670	1	SYN1_MOUSE	O88935 mus musculu
40	155.5	4.1	924	1	104K_THEPA	P15711 theileria p
41	154.5	4.1	568	1	RGS3_MOUSE	Q9dc04 mus musculu
42	154.5	4.1	1174	1	KPC1_COCHE	O42632 cochliobolu
43	154.5	4.1	1443	1	SYJ2_HUMAN	O15056 homo sapien
44	154	4.1	733	1	VINE_MOUSE	Q9r1z8 mus musculu
45	154	4.1	777	1	RTN1_RAT	Q64548 rattus norv

ALIGNMENTS

RESULT 1
JIP1_HUMAN
ID JIP1_HUMAN STANDARD; PRT; 711 AA.
AC Q9UQF2; O43407;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE C-jun-amino-terminal kinase interacting protein 1 (JNK-interacting
DE protein 1) (JIP-1) (JNK MAP kinase scaffold protein 1) (Islet-brain-1)
DE (IB-1) (Mitogen-activated protein kinase 8-interacting protein 1).
GN MAPK8IP1 OR JIP1 OR IB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Insulinoma;
RX MEDLINE=99134298; PubMed=9933567;
RA Mooser V., Maillard A., Bonny C., Steinmann M., Shaw P., Yarnall D.P.,
RA Burns D.K., Schorderet D.F., Nicod P., Waeber G.;
RT "Genomic organization, fine-mapping, and expression of the human
RT islet-brain 1 (IB1)/C-jun-amino-terminal kinase interacting protein-1
RT (JIP-1) gene.";
RL Genomics 55:202-208(1999).
RN [2]
RP SEQUENCE OF 468-711 FROM N.A.
RC TISSUE=Brain;
RA Yu W., Sarginson J., Gibbs R.A.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP RHOGEF BINDING, AND SUBCELLULAR LOCATION.
RX MEDLINE=20044776; PubMed=10574993;
RA Meyer D., Liu A., Margolis B.;
RT "Interaction of c-Jun amino-terminal kinase interacting protein-1 with
RT p190 rhoGEF and its localization in differentiated neurons.";
RL J. Biol. Chem. 274:35113-35118(1999).
RN [4]
RP MUTAGENESIS, AND KINESIN BINDING.
RX MEDLINE=21135887; PubMed=11238452;
RA Verhey K.J., Meyer D., Deehan R., Blenis J., Schnapp B.J.,
RA Rapoport T.A., Margolis B.;
RT "Cargo of kinesin identified as JIP scaffolding proteins and
RT associated signaling molecules.";
RL J. Cell Biol. 152:950-970(2001).
RN [5]
RP VARIANT NIDDM ASN-59.
RX MEDLINE=20164330; PubMed=10700186;
RA Waeber G., Delplanque J., Bonny C., Mooser V., Steinmann M.,
RA Widmann C., Maillard A., Miklossy J., Dina C., Hani E.H., Vionnet N.,
RA Nicod P., Boutin P., Froguel P.;
RT "The gene, MAPK8IP1, encoding islet-brain-1, is a candidate for type 2
RT diabetes.";
RL Nat. Genet. 24:291-295(2000).
CC -!- FUNCTION: The JNK-interacting protein (JIP) group of scaffold
CC proteins selectively mediates JNK signaling by aggregating

```
CC       specific components of the MAPK cascade to form a functional JNK
CC       signaling module. JIP1 is required for JNK activation in response
CC       to excitotoxic stress (By similarity). Cytoplasmic JIP1 causes
CC       inhibition of JNK-regulated activity by retaining JNK in the
CC       cytoplasm and by inhibiting the JNK phosphorylation of c-Jun (By
CC       similarity). May also participate in ApoER2-specific reelin
CC       signaling. Directly, or indirectly, regulates GLUT2 gene
CC       expression and beta-cell function (By similarity). Appears to have
CC       a role in cell signaling in mature and developing nerve terminals.
CC   -!- SUBUNIT: Forms homo-or heterooligomeric complexes. Binds specific
CC       components of the JNK signaling pathway namely JNK, MAPKK7 and
CC       MLK2, MLK3 and DLK (By similarity). Also binds the proline-rich
CC       domain-containing splice variant of apolipoprotein E receptor 2
CC       (ApoER2) (By similarity). Binds the TPR motif-containing C-
CC       terminal of kinesin light chain. Interacts, via the PID domain,
CC       with rhoGEF. Binds the cytoplasmic tails of LRP1 and LRP2
CC       (Megalin) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic. Accumulates in cell surface
CC       projections (By similarity). Under certain stress conditions,
CC       translocates to the perinuclear region of neurons (By similarity).
CC       In insulin-secreting cells, detected in both the cytoplasm and
CC       nucleus (By similarity).
CC   -!- TISSUE SPECIFICITY: Highly expressed in brain. Expressed in
CC       neurons, localizing to neurite tips in differentiating cells. Also
CC       expressed in the pancreas, testis and prostate. Low levels in
CC       heart, ovary and small intestine.
CC   -!- PTM: Phosphorylated by JNK in differentiated cells.
CC   -!- DISEASE: Defects in MAPK8IP1 may contribute to autosomal dominant
CC       non-insulin-dependent diabetes mellitus type II (NIDDM), which is
CC       characterized by an autosomal dominant mode of inheritance, onset
CC       during childhood (usually before 25 years of age) and a primary
CC       defect in insulin secretion.
CC   -!- SIMILARITY: BELONGS TO THE JIP SCAFFOLD PROTEINS FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PID DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF074091; AAD20443.1; -.
DR   EMBL; AF007134; AAC19150.1; -.
DR   TRANSFAC; T04465; -.
DR   Genew; HGNC:6882; MAPK8IP1.
DR   MIM; 604641; -.
DR   MIM; 125853; -.
DR   InterPro; IPR000050; PID_domain.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00018; SH3; 2.
DR   Pfam; PF00640; PID; 2.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00462; PTB; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS01179; PID; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   SH3 domain; Phosphorylation; Disease mutation; Diabetes mellitus.
FT   DOMAIN       42     48       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       79     84       POLY-GLY.
FT   DOMAIN      107    116       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      127    285       JNK-BINDING DOMAIN (JBD).
FT   DOMAIN      331    334       POLY-GLU (ACIDIC).
FT   DOMAIN      359    363       POLY-PRO.
FT   DOMAIN      488    549       SH3.
FT   DOMAIN      561    700       PID.
FT   VARIANT      59     59       S -> N (IN NIDDM).
FT                                /FTId=VAR_012243.
FT   MUTAGEN     704    704       P->A: KINESIN BINDING.
FT   MUTAGEN     709    709       Y->A: NO KINESIN BINDING.
SQ   SEQUENCE   711 AA;  77524 MW;  55EA53B30080A751 CRC64;
```

```
  Query Match           99.9%;  Score 3789;  DB 1;  Length 711;
  Best Local Similarity  99.9%;  Pred. No. 3.7e-184;
  Matches  710;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60

Qy      61 QCKDTLSLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAAR 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 QCKDTLSLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAAR 120

Qy     121 RPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNNN 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 RPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNNN 180

Qy     181 SLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRRS 240
           ||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||
Db     181 SLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSDELPPQSGPAPTTDRGTSTDSPCRRS 240

Qy     241 TATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAFL 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAFL 300

Qy     301 PPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEPP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 PPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEPP 360

Qy     361 PPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASVS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 PPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASVS 420

Qy     421 SPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLF 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 SPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLF 480

Qy     481 SCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFP 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 SCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFP 540

Qy     541 AYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRLT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 AYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRLT 600

Qy     601 VHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGFI 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 VHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGFI 660

Qy     661 TKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
```

```
RESULT 2
JIP1_MOUSE
ID   JIP1_MOUSE     STANDARD;      PRT;   707 AA.
AC   Q9WVI9; Q9WVI8; Q9WVI7; Q9R1H9; Q9R1Z1; O35145;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   C-jun-amino-terminal kinase interacting protein 1 (JNK-interacting
DE   protein 1) (JIP-1) (JNK MAP kinase scaffold protein 1) (Islet-brain-1)
DE   (IB-1) (Mitogen-activated protein kinase 8-interacting protein 1).
GN   MAPK8IP1 OR PRKM8IP OR JIP1 OR IB1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM JIP-1A), AND POSSIBLE FUNCTION.
RC   TISSUE=Brain;
RX   MEDLINE=97382313; PubMed=9235893;
```

```
RA   Dickens M., Rogers J.S., Cavanagh J., Raitano A., Xia Z.,
RA   Halpern J.R., Greenberg M.E., Sawyers C.L., Davis R.J.;
RT   "A cytoplasmic inhibitor of the JNK signal transduction pathway.";
RL   Science 277:693-696(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS JIP-1B; JIP-1C; JIP-1D AND JIP-1E).
RC   STRAIN=BALB/c; TISSUE=Brain;
RX   MEDLINE=99196470; PubMed=10098834;
RA   Kim I.-J., Lee K.-W., Park B.Y., Lee J.-K., Park J., Choi I.Y.,
RA   Eom S.-J., Chang T.-S., Kim M.J., Yeom Y.I., Chang S.K., Lee Y.-D.,
RA   Choi E.-J., Han P.-L.;
RT   "Molecular cloning of multiple splicing variants of JIP-1
RT   preferentially expressed in brain.";
RL   J. Neurochem. 72:1335-1343(1999).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM JIP-1B), AND CHARACTERIZATION.
RC   TISSUE=Brain;
RX   MEDLINE=99422004; PubMed=10490659;
RA   Yasuda J., Whitmarsh A.J., Cavanagh J., Sharma M., Davis R.J.;
RT   "The JIP group of mitogen-activated protein kinase scaffold
RT   proteins.";
RL   Mol. Cell. Biol. 19:7245-7254(1999).
RN   [4]
RP   TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.
RX   MEDLINE=20177638; PubMed=10712642;
RA   Pellet J.-B., Haefliger J.-A., Staple J.K., Widmann C., Welker E.,
RA   Hirling H., Bonny C., Nicod P., Catsicas S., Waeber G., Riederer B.M.;
RT   "Spatial, temporal and subcellular localization of islet-brain 1
RT   (IB1), a homologue of JIP-1, in mouse brain.";
RL   Eur. J. Neurosci. 12:621-632(2000).
RN   [5]
RP   INTERACTION WITH LRPS.
RX   MEDLINE=20400498; PubMed=10827173;
RA   Gotthardt M., Trommsdorff M., Nevitt M.F., Shelton J.,
RA   Richardson J.A., Stockinger W., Nimpf J., Herz J.;
RT   "Interactions of the low density lipoprotein receptor gene family with
RT   cytosolic adaptor and scaffold proteins suggest diverse biological
RT   functions in cellular communication and signal transduction.";
RL   J. Biol. Chem. 275:25616-25624(2000).
RN   [6]
RP   FUNCTION.
RX   MEDLINE=21446505; PubMed=11562351;
RA   Whitmarsh A.J., Kuan C.-Y., Kennedy N.J., Kelkar N., Haydar T.F.,
RA   Mordes J.P., Appel M., Rossini A.A., Jones S.N., Flavell R.A.,
RA   Rakic P., Davis R.J.;
RT   "Requirement of the JIP1 scaffold protein for stress-induced JNK
RT   activation.";
RL   Genes Dev. 15:2421-2432(2001).
CC   -!- FUNCTION: The JNK-interacting protein (JIP) group of scaffold
CC       proteins selectively mediates JNK signaling by aggregating
CC       specific components of the MAPK cascade to form a functional JNK
CC       signaling module. JIP1 is required for JNK activation in response
CC       to excitotoxic stress. Cytoplasmic JIP1 causes inhibition of JNK-
CC       regulated activity by retaining JNK in the cytoplasm and by
CC       inhibiting the JNK phosphorylation of c-Jun. May also participate
CC       in ApoER2-specific reelin signaling. Directly, or indirectly,
CC       regulates GLUT2 gene expression and beta-cell function (By
CC       similarity). Appears to have a role in cell signaling in mature
CC       and developing nerve terminals.
CC   -!- SUBUNIT: Forms homo-or heterooligomeric complexes. Binds specific
CC       components of the JNK signaling pathway namely JNK, MAPKK7 and
CC       MLK2, MLK3 and DLK. Also binds the proline-rich domain-containing
CC       splice variant of apolipoprotein E receptor 2 (ApoER2). Binds the
CC       TPR motif-containing C-terminal of kinesin light chain. Interacts,
CC       via the PID domain, with rhoGEF (By similarity). Binds the
CC       cytoplasmic tails of LRP1 and LRP2 (Megalin).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic. Accumulates in cell surface
CC       projections. Under certain stress conditions, translocates to the
CC       perinuclear region of neurons. In insulin-secreting cells,
CC       detected in both the cytoplasm and nucleus (By similarity).
CC   -!- ALTERNATIVE PRODUCTS: 5 isoforms; JIP-1a/1, JIP-1b (shown here),
CC       JIP-1c/2a, JIP-1d/2B and JIP-1e/3; are produced by alternative
CC       splicing.
CC   -!- TISSUE SPECIFICITY: Expressed predominantly in the brain and
CC       insulin-secreting cells. In the brain, high expression found in
CC       the cerebral cortex and hippocampus. Localizes in the synaptic
CC       regions of the olfactory bulb, retina, cerebral and cerebellar
CC       cortex and hippocampus. Also expressed in a restricted number of
CC       axons, including mossy fibers from the hippocampal dentate gyrus,
CC       soma, dendrites and axons of cerebellar Purkinje cells. Also
CC       expressed in kidney, testis and prostate. Low levels in heart,
CC       ovary and small intestine. Isoform JIP-1b is more predominant in
CC       the brain than isoform JIP-1a. Jip1-a is expressed both in the
CC       brain and kidney, isoforms JIP-1c, JIP-1d and JIP-1e are brain
CC       specific.
CC   -!- DEVELOPMENTAL STAGE: Low levels at prenatal stage E15, increased
CC       levels during the first postnatal days, with a plateau at
CC       postnatal day 15.
CC   -!- INDUCTION: Upon neuron differentiation.
CC   -!- PTM: Phosphorylated by JNK in differentiated cells.
CC   -!- SIMILARITY: BELONGS TO THE JIP SCAFFOLD PROTEINS FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PID DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF003115; AAB66317.1; -.
DR   EMBL; AF109768; AAD38346.1; -.
DR   EMBL; AF109769; AAD38347.1; -.
DR   EMBL; AF109770; AAD38348.1; -.
DR   EMBL; AF109771; AAD38349.1; -.
DR   EMBL; AF054611; AAD22580.1; -.
DR   TRANSFAC; T04468; -.
DR   MGD; MGI:1309464; Prkm8ip.
DR   InterPro; IPR000050; PID_domain.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00018; SH3; 5.
DR   Pfam; PF00640; PID; 5.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS01179; PID; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   SH3 domain; Alternative splicing; Phosphorylation.
FT   DOMAIN       41     47       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      107    116       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      127    281       JNK-BINDING DOMAIN (JBD).
FT   DOMAIN      355    359       POLY-PRO.
FT   DOMAIN      484    545       SH3.
FT   DOMAIN      557    696       PID.
FT   VARSPLIC      1     33       MAERESGLGGGAASPPAASPFLGLHIASPPNFR -> MQLV
FT                                LKMDSSPDNDSWLEDQWEHW (IN ISOFORM JIP-1C
FT                                AND ISOFORM JIP-1D).
FT   VARSPLIC      1     90       MISSING (IN ISOFORM JIP-1E).
FT   VARSPLIC     69     93       MISSING (IN ISOFORM JIP-1D).
FT   VARSPLIC    558    604       MISSING (IN ISOFORM JIP-1A).
FT   CONFLICT    144    145       PG -> A (IN REF. 2).
FT   CONFLICT    593    593       R -> RP (IN REF. 2).
SQ   SEQUENCE   707 AA;  77281 MW;  274013B12D91049D CRC64;

  Query Match             93.6%;  Score 3549;  DB 1;  Length 707;
  Best Local Similarity   93.7%;  Pred. No. 4.4e-172;
  Matches  667;  Conservative   13;  Mismatches   26;  Indels    6;  Gaps    3;

Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAERES-GLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 59

Qy      61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
           |||||||||||||||||||  | |||||||||||||||| ||||||||||::|:| ||
Db      60 QCKDTLSLRPPRAGLLSAGSSGSAGSRLQAEMLQMDLIDAAGDTPGAEDDEEEEDDELAA 119
```

```
Qy     120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
           :||| ||||||| |:|| |    ||||  | ||||||||||||||||||||||||||||||
Db     120 QRPGVGPPKAESNQDPAPR----SQGQGPGTGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 175

Qy     180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
           ||||||||||||||||||||||||||||||||||:|||||  | || ||||||||||||
Db     176 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSDELPPQGSPVPTQDRGTSTDSPCRR 235

Qy     240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
           | ||||||| |||| ||||||||||||||||||||||||||||||||||| ||||| |
Db     236 SAATQMAPPSGPPATAPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDPAEPTSTF 295

Qy     300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
           :|||||||||||||||||  ||||||||||||:|||||||||||||||||||||||||||
Db     296 MPPTESRMSVSSDPDPAAYSVTAGRPHPSISEEDEGFDCLSSPERAEPPGGGWRGSLGEP 355

Qy     360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     356 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASA 415

Qy     420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
           ||||||||||||||||:|:|| ||||||||||||||||||||||||||||||||||||||
Db     416 SSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 475

Qy     480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
           |||:||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     476 FSCVINGEEHEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 535

Qy     540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
           ||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db     536 PAYYAIEVTKEPEHMAALAKNSDWIDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 595

Qy     600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
           |||||||||||||||||||||||||: ||||||||||||||||||||||||||||||||
Db     596 TVHFNPPSSCVLEISVRGVKIGVKADDALEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 655

Qy     660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     656 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 707
```

RESULT 3
JIP1_RAT
ID JIP1_RAT STANDARD; PRT; 708 AA.
AC Q9R237; O88979; Q9R1H8; Q9WVI6; Q9WVI5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE C-jun-amino-terminal kinase interacting protein 1 (JNK-interacting
DE protein 1) (JIP-1) (JNK MAP kinase scaffold protein 1) (Islet-brain-1)
DE (IB-1) (Mitogen-activated protein kinase 8-interacting protein 1)
DE (JIP-1 related protein) (JRP).
GN MAPK8IP1 OR JIP1 OR IB1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Pancreas;
RX MEDLINE=98113135; PubMed=9442013;
RA Bonny C., Nicod P., Waeber G.;
RT "IB1, a JIP-1-related nuclear protein present in insulin-secreting
RT cells.";
RL J. Biol. Chem. 273:1843-1846(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC STRAIN=Fischer; TISSUE=Fibroblast;
RA Chen Y., Talmage D.;
RT "JIP-1 related protein (JRP).";
RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=99196470; PubMed=10098834;
RA Kim I.-J., Lee K.-W., Park B.Y., Lee J.-K., Park J., Choi I.Y.,
RA Eom S.-J., Chang T.-S., Kim M.J., Yeom Y.I., Chang S.K., Lee Y.-D.,
RA Choi E.-J., Han P.-L.;
RT "Molecular cloning of multiple splicing variants of JIP-1
RT preferentially expressed in brain.";
RL J. Neurochem. 72:1335-1343(1999).
CC -!- FUNCTION: The JNK-interacting protein (JIP) group of scaffold
CC proteins selectively mediates JNK signaling by aggregating
CC specific components of the MAPK cascade to form a functional JNK
CC signaling module. JIP1 is required for JNK activation in response
CC to excitotoxic stress (By similarity). Cytoplasmic JIP1 causes
CC inhibition of JNK-regulated activity by retaining JNK in the
CC cytoplasm and by inhibiting the JNK phosphorylation of c-Jun (By
CC similarity). May also participate in ApoER2-specific reelin
CC signaling. Directly, or indirectly, regulates GLUT2 gene
CC expression and beta-cell function. Appears to have a role in cell
CC signaling in mature and developing nerve terminals.
CC -!- SUBUNIT: Forms homo-or heterooligomeric complexes. Binds specific
CC components of the JNK signaling pathway namely JNK, MAPKK7 and
CC MLK2, MLK3 and DLK (By similarity). Also binds the proline-rich
CC domain-containing splice variant of apolipoprotein E receptor 2
CC (ApoER2) (By similarity). Binds the TPR motif-containing C-
CC terminal of kinesin light chain (By similarity). Interacts, via
CC the PID domain, with rhoGEF (By similarity). Binds the cytoplasmic
CC tails of LRP1 and LRP2 (Megalin) (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Accumulates in cell surface
CC projections (By similarity). Under certain stress conditions,
CC translocates to the perinuclear region of neurons (By similarity).
CC In insulin-secreting cells, detected in both the cytoplasm and
CC nucleus.
CC -!- ALTERNATIVE PRODUCTS: 3 isoforms; 1/JIP-1a/JIP-1b (shown here),
CC 2/JIP-1c/2A and 3/JIP-1d; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Highly expressed in brain and pancreatic beta-
CC cells. Weaker expression found in kidney.
CC -!- PTM: Phosphorylated by JNK in differentiated cells (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE JIP SCAFFOLD PROTEINS FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PID DOMAIN.
CC ---

CC ---
DR EMBL; AF108959; AAD22543.1; ALT_INIT.
DR EMBL; AF092450; AAC62110.1; -.
DR EMBL; AF109772; AAD38350.1; -.
DR EMBL; AF109773; AAD38351.1; ALT_INIT.
DR EMBL; AF109774; AAD38352.1; -.
DR TRANSFAC; T04466; -.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 4.
DR Pfam; PF00640; PID; 4.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00462; PTB; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS01179; PID; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain; Alternative splicing; Phosphorylation.
FT DOMAIN 41 47 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 107 115 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 126 282 JNK-BINDING DOMAIN (JBD).
FT DOMAIN 356 360 POLY-PRO.
FT DOMAIN 485 546 SH3.
FT DOMAIN 558 697 PID.

```
FT   VARSPLIC      1     33       MAERESGLSGGAASPPAASPFLGLHIASPPNFR -> MQLV
FT                                LKMDSSPDNDSWLEDQWERW (IN ISOFORM 2).
FT   VARSPLIC     69     93       MISSING (IN ISOFORM 3).
FT   VARSPLIC    708    708       E -> EPMAQVQLQVDLEIKRAAAEQKLISEEDLNGAA
FT                                (IN ISOFORM 3).
FT   CONFLICT     38     38       I -> V (IN REF. 2).
FT   CONFLICT    149    149       S -> C (IN REF. 1).
SQ   SEQUENCE   708 AA;  77317 MW;  4923FD55F1C511F4 CRC64;

  Query Match             93.5%;  Score 3544.5;  DB 1;  Length 708;
  Best Local Similarity   94.0%;  Pred. No. 7.5e-172;
  Matches  669;  Conservative   11;  Mismatches   27;  Indels    5;  Gaps    4;

Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
           |||||| || |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAERES-GLSGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 59

Qy      61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
           |||||||||||||||||||  | ||||||||||||||||  |||||||||||| | ||
Db      60 QCKDTLSLRPPRAGLLSAGSSGSAGSRLQAEMLQMDLIDAASDTPGAEDDEEDDD-ELAA 118

Qy     120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
           :||| || |||||||||| ||  |||  | ||||||||||||||||||||||||||||||
Db     119 QRPGVGPSKAESGQEPASRSQG--QGQGPGTGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 176

Qy     180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
           ||||||||||||||||||||||||||||||:|||||  | || ||||||||||||
Db     177 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSDELPPQGSPVPTQDRGTSTDSPCRR 236

Qy     240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
           : ||||||| ||||  |||||||||||||||||||||||||||||||||| ||||| |
Db     237 TAATQMAPPSGPPATAPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDPAEPTSTF 296

Qy     300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
           |||||||||||||||  ||||||||||:||||||||:|||||||||||||||||||||||
Db     297 LPPTESRMSVSSDPDPAAYSVTAGRPHPSISEEDEGFDCLSSPEQAEPPGGGWRGSLGEP 356

Qy     360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     357 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASA 416

Qy     420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
           |||||||||||||:|:|{ ||||||||||||||||||||||||||||||||||||||||
Db     417 SSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 476

Qy     480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
           |||:|||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     477 FSCVINGEEHEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 536

Qy     540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
           ||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db     537 PAYYAIEVTKEPEHMAALAKNSDWIDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 596

Qy     600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
           |||||||||||||||||||||::|||||||||||||||||||||||||||||||||||||
Db     597 TVHFNPPSSCVLEISVRGVKIGVKADEAQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 656

Qy     660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     657 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 708


RESULT 4
JIP2_MOUSE
ID   JIP2_MOUSE     STANDARD;      PRT;   830 AA.
AC   Q9ERE9; Q9CXI4;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   C-jun-amino-terminal kinase interacting protein 2 (JNK-interacting
DE   protein 2) (JIP-2) (JNK MAP kinase scaffold protein 2) (Islet-brain-2)
DE   (IB-2) (Mitogen-activated protein kinase 8-interacting protein 2).
GN   MAPK8IP2 OR JIP2 OR IB2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND INTERACTION WITH APOER2.
RC   STRAIN=BALB/c; TISSUE=Brain;
RX   MEDLINE=20400499; PubMed=10827199;
RA   Stockinger W., Brandes C., Fasching D., Hermann M., Gotthardt M.,
RA   Herz J., Schneider W.J., Nimpf J.;
RT   "The reelin receptor ApoER2 recruits JNK-interacting proteins-1 and
RT   -2.";
RL   J. Biol. Chem. 275:25625-25632(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- FUNCTION: The JNK-interacting protein (JIP) group of scaffold
CC       proteins selectively mediates JNK signaling by aggregating
CC       specific components of the MAPK cascade to form a functional JNK
CC       signaling module. JIP2 inhibits IL1 beta-induced apoptosis in
CC       insulin-secreting cells (By similarity).
CC   -!- SUBUNIT: Forms homo-or heterooligomeric complexes. Binds specific
CC       components of the JNK signaling pathway namely JNK, MAPKK7 and
CC       MLK2, MLK3 and DLK (By similarity). Also binds the proline-rich
CC       domain-containing splice variant of apolipoprotein E receptor 2
CC       (ApoER2). Binds the TPR motif-containing C-terminal of kinesin
CC       light chain. Binds the cytoplasmic tails of LRP1 and LRP2
CC       (Megalin).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic. Accumulates in cell surface
CC       projections (By similarity).
CC   -!- TISSUE SPECIFICITY: Highly expressed in brain. Expressed in all
CC       neurons. Also expressed in testis, primarily in the epididymal
CC       epidermis.
CC   -!- INDUCTION: Upon neuron differentiation.
CC   -!- SIMILARITY: BELONGS TO THE JIP SCAFFOLD PROTEINS FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PID DOMAIN.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to a stop
CC       codon in position 701.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF310135; AAG31800.1; -.
DR   EMBL; AK014339; BAB29282.1; ALT_SEQ.
DR   MGD; MGI:1926555; Mapk8ip2.
DR   InterPro; IPR000050; PID_domain.
```

```
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00640; PID; 1.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00462; PTB; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS01179; PID; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   SH3 domain; Alternative splicing.
FT   DOMAIN       30     36       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       85    104       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      111    278       JNK-BINDING DOMAIN (JBD).
FT   DOMAIN      154    157       POLY-ASN.
FT   DOMAIN      282    293       SER-RICH.
FT   DOMAIN      420    437       PRO-RICH.
FT   DOMAIN      472    487       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      610    671       SH3.
FT   DOMAIN      683    819       PID.
FT   CONFLICT    216    216       P -> Q (IN REF. 2).
SQ   SEQUENCE   830 AA;  89899 MW;  7EC8EAD19A90163C CRC64;

  Query Match             34.1%;  Score 1292.5;  DB 1;  Length 830;
  Best Local Similarity   36.9%;  Pred. No. 2.2e-58;
  Matches  328;  Conservative   94;  Mismatches  192;  Indels  275;  Gaps   31;

Qy      26 HIASPPNFRLTHDISLEEFEDEDLSEITDECGISL-----QC-KDTLSLRPPRAGLLSAG 79
           |  |||  |   |||||||:|||||||||:||: |      | ||:|||
Db      14 HSLSPPGCRPPQDISLEEFDDEDLSEITDDCGLGLSYDSDHCEKDSLSL---------- 62

Qy      80 GGGAGSR------LQAEMLQMDLIDATGDTPGAEDDEEDDDEERAARRPGAGPPKAESGQ 133
             |  :       | :  : ::||       |:||:|||||          : |  :
Db      63 --GRSEQPHPICSFQDDFQEFEMID---------DNEEEDDEEE----------EEEEEE 101

Qy     134 EPASRGQGQSQGQSQGPGS----GDTYRPK---------RPTTLNLFPQVPRSQDTLNNN 180
           |     ||:: |   ||||    ||:  |          |||||:|      :||:||||
Db     102 EEDGDRQGKAGG---GPGSQALAGDSLIPSPSLEESHKLRPTTLHL--TTLGAQDSLNNN 156

Qy     181 SLG----KKHSWQDRVSRS--SSPLK-------------------------TGEQTPPH 208
           : |        |||: | ||    |||                          | | |
Db     157 NGGFTSAPPSSWQETVLRSPAQEPLKELPAPLLPAEEERHEVQSLARPGCDCEGNQPPE- 215

Qy     209 EHICLSEELPPQS--GPAPTTDRGTSTDSPCR-------RSTATQMAPPGGPPAAPPGGR 259
                 || |  | :|::| |   |           | :: ::: ||       | |
Db     216 ---------PPASSGGASPSSDPGIEADLRSHSSGGHEGRRSSQELSSPGSDSEDAGGAR 266

Qy     260 ------------------GHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEP------ 295
                           | |   |  :  :   ||:     :|   |    ||
Db     267 LGRMISSISETELELSSDGGSSSGRSSHLTNSIEEAS-----SPASEPEPEPEPLHEPPR 321

Qy     296 TSAFLP------------PTESRMSVSSDPD------------------PAAYPSTAGRP 325
             ||||              :||   :| | |                |  |    |
Db     322 RPAFLPVGQDDTNSEYESGSESEPDLSEDADSPWLLSNLVSRMISEGSSPIRCPGQCLSP 381

Qy     326 HPSISEEEEGFDCLSSPERAEPPGGGWRGSL-------GEPPPPPRAS------------ 366
           | : ||       | |:  : |  |    |         | ||| | |
Db     382 APRLPEEAAS-QANSVPQDCQDPEAGPHVELVDMDTLCGPPPPAPAAPRLGPAQPGPCLF 440

Qy     367 LSSDT---------------------------------------------------- 371
           ||: |
Db     441 LSNPTRDTITPLWATPGRTARPGRSCSAACSEEEEEDEEEDEEDEEDAEDSVVPPGSRTT 500

Qy     372 -------SALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSD-ESDSATVYDNC-ASVSSP 422
                ::| ||:|||||||||| |||||||| | |  :| | ||     :| ||
Db     501 GSTAPLDASLVYDAVKYTLVVDEHTQLELVSLRRCAGLGNDSEEDS-----SCEASEEEA 555

Qy     423 YESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLFSC 482
             : :| :       | |    ||::|| ||: ||||||||::  |||| ||||||||
Db     556 GATLLGSD-------QVP----EDASPDSPDLTFSKKFLNVFVNSTSRSSSTESFGLFSC 604

Qy     483 IINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFPAY 542
           ::||||:||||||:|||:||| ||||:||||:||| : :|:|:  :|||| ||||||:
Db     605 VVNGEERFQTHRAVFRFIPRHPDELELDVDDPVLVEAEEDDFWFRGFNMRTGERGVFPAF 664

Qy     543 YAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRLTVH 602
           ||  |    : :    ::  |||:| |:|||||:|| |:|| :|||||||||| |:|||
Db     665 YAHAVPGPAKDLLGSKRSPCWVDRFDVQFLGSVEVPCHQGNGILCAAMQKIATARKLTVH 724

Qy     603 FNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGFITK 662
             ||:|| ||||:|||| :     :     :|||||:|||||| ||:|: ||||||
Db     725 LRPPASCDLEISLRGVKLSLSGGGPE---FQRCSHFFQMKNISFCGCHPRNSCYFGFITK 781

Qy     663 HPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
           ||   |||||||||::| : :| |||||| ::|:: : : |||||||||
Db     782 HPLLSRFACHVFVSQESMRPVARSVGRAFLEYYQEHLAFACPTEDIYLE 830


RESULT 5
JIP2_HUMAN
ID   JIP2_HUMAN     STANDARD;      PRT;   824 AA.
AC   Q13387; Q9UKQ4; Q9NZ59; Q99771;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   C-jun-amino-terminal kinase interacting protein 2 (JNK-interacting
DE   protein 2) (JIP-2) (JNK MAP kinase scaffold protein 2) (Islet-brain-2)
DE   (IB-2) (Mitogen-activated protein kinase 8-interacting protein 2).
GN   MAPK8IP2 OR JIP2 OR IB2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain;
RX   MEDLINE=99422004; PubMed=10490659;
RA   Yasuda J., Whitmarsh A.J., Cavanagh J., Sharma M., Davis R.J.;
RT   "The JIP group of mitogen-activated protein kinase scaffold
RT   proteins.";
RL   Mol. Cell. Biol. 19:7245-7254(1999).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain, and Insulinoma;
RX   MEDLINE=20221377; PubMed=10756100;
RA   Negri S., Oberson A., Steinmann M., Sauser C., Nicod P., Waeber G.,
RA   Schorderet D.F., Bonny C.;
RT   "cDNA cloning and mapping of a novel islet-brain/JNK interacting
RT   protein.";
RL   Genomics 64:324-330(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20057165; PubMed=10591208;
RA   Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA   Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA   Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA   Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA   Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA   Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA   Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA   Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA   Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA   Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA   Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,
RA   Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA   Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA   Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA   Mcclay J., Mclaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA   Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA   Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA   Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA   Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA   Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA   Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA   Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA   Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA   Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA   Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
```

```
RA   Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA   Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA   Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA   Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA   Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA   Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA   Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA   Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA   Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
RA   Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA   Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA   McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
RA   Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,
RA   Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
RA   Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA   Tilahun Y., Wright H.;
RT   "The DNA sequence of human chromosome 22.";
RL   Nature 402:489-495(1999).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RA   Adams M.D.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 599-824 FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain;
RA   Yu W., Gibbs R.A.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: The JNK-interacting protein (JIP) group of scaffold
CC       proteins selectively mediates JNK signaling by aggregating
CC       specific components of the MAPK cascade to form a functional JNK
CC       signaling module. JIP2 inhibits IL1 beta-induced apoptosis in
CC       insulin-secreting cells.
CC   -!- SUBUNIT: Forms homo-or heterooligomeric complexes. Binds specific
CC       components of the JNK signaling pathway namely JNK, MAPKK7 and
CC       MLK2, MLK3 and DLK. Also binds the proline-rich domain-containing
CC       splice variant of apolipoprotein E receptor 2 (ApoER2) (By
CC       similarity). Binds the TPR motif-containing C-terminal of kinesin
CC       light chain (By similarity). Binds the cytoplasmic tails of
CC       LRP1 and LRP2 (Megalin) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic. Accumulates in cell surface
CC       projections.
CC   -!- ALTERNATIVE PRODUCTS: 3 isoforms; 1 (shown here), 2 and 3; may be
CC       produced by alternative splicing. Isoform 3 might be artefactual
CC       as it is only predicted from a genomic sequence.
CC   -!- TISSUE SPECIFICITY: Expressed mainly in the brain and pancreas,
CC       including insulin-secreting cells. In the nervous system, more
CC       abundantly expressed in the cerebellum, pituitary gland, occipital
CC       lobe and the amygdala. Also expressed in fetal brain. Very low
CC       levels found in uterus, ovary, prostate, colon, testis, adrenal
CC       gland, thyroid gland and salivary gland.
CC   -!- SIMILARITY: BELONGS TO THE JIP SCAFFOLD PROTEINS FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PID DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF136382; AAF00980.1; -.
DR   EMBL; AF218778; AAF32323.1; -.
DR   EMBL; AL021708; CAA16714.1; ALT_SEQ.
DR   EMBL; U62317; AAB03340.1; -.
DR   EMBL; U79261; AAB50207.1; -.
DR   Genew; HGNC:6883; MAPK8IP2.
DR   InterPro; IPR000050; PID_domain.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00018; SH3; 4.
DR   Pfam; PF00640; PID; 4.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00462; PTB; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS01179; PID; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   SH3 domain; Alternative splicing.
FT   DOMAIN       30     36       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       85    103       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      110    275       JNK-BINDING DOMAIN (JBD).
FT   DOMAIN      279    291       SER-RICH.
FT   DOMAIN      417    434       PRO-RICH.
FT   DOMAIN      469    481       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      604    665       SH3.
FT   DOMAIN      677    813       PID.
FT   VARSPLIC      1     57       MADRAEMFSLSTFHSLSPPGCRPPQDISLEEFDDEDLSEIT
FT                                DDCGLGLSYDSDHCEK -> MLPDFPSPSTWAPGLLLPSGP
FT                                ALLSPSVLQ (IN ISOFORM 2 AND ISOFORM 3).
FT   VARSPLIC    152    394       MISSING (IN ISOFORM 3).
FT   VARSPLIC    415    439       MISSING (IN ISOFORM 3).
FT   VARSPLIC    768    768       C -> CEAPQGAAFQWERGVDRKRVLQTRGNVQPHLGAGQ
FT                                GAALNRATEGSSTGSEKGEWTPLVIMELTQSVNSC (IN
FT                                ISOFORM 3).
SQ   SEQUENCE   824 AA;  87974 MW;  7EFC8F3BC58E37BB CRC64;
```

Query Match 33.8%; Score 1280.5; DB 1; Length 824;
Best Local Similarity 36.6%; Pred. No. 8.9e-58;
Matches 326; Conservative 91; Mismatches 190; Indels 283; Gaps 30;

```
Qy      26 HIASPPNFRLTHDISLEEFEDEDLSEITDECGISL-----QC-KDTLSLRPPRAGLLSAG 79
           |  |||  |   |||||||:||||||||:||: |      | ||:|||
Db      14 HSLSPPGCRPPQDISLEEFDDEDLSEITDDCGLGLSYDSDHCEKDSLSL---------- 62

Qy      80 GGGAGSR------LQAEMLQMDLIDATGDTPGAEDDEEDDDEERAARRPGAGPPKAESGQ 133
             |   :       | :  : ::||         |:||:|||:          : |  :
Db      63 --GRSEQPHPICSFQDDFQEFEMID---------DNEEEDDED----------EEEEEE 100

Qy     134 EPASRGQGQSQGQ--SQGPGSG--------DTYRPKRPTTLNLFPQVPRSQDTLNNN--- 180
           |    |:||  |   |: |  |         :     ||||| |      :||:||||
Db     101 EEEGDGEGQEGGDPGSEAPAPGPLIPSPSVEEPHKHRPTTLRL--TTLGAQDSLNNNGGF 158

Qy     181 SLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGT-STDSPCR- 238
           | :  |||:    | :|                | |    || | || |     || |
Db     159 DLVRPASWQETALCSPAP---------------EALRELPGPLPATDTGPGGAQSPVRP 202

Qy     239 ------RSTATQMAPPGGPPAAPP----------GGRGHSHRDR--------------- 266
                  |   || |  |:: |             ||||     :
Db     203 GCDCEGNRPAEPPAPGGTSPSSDPGIEADLRSRSSGGRGGRRSSQELSSPGSDSEDAGGA 262

Qy     267 -----IHYQADVRLEATEE---------IYLT--------PVQRPPDAAEP--TSAFLP- 301
                 |   ::  || : :          :||        |   |    ||    ||||
Db     263 RLGRMISSISETELELSSDGGSSSSGRSSHLTNSIEEASSPASEPEPPREPPRRPAFLPV 322

Qy     302 ----------PTESRMSVSSDPD---------------------PAAYPSTAGRP--- 325
                    :||   :| | |                       |    | | ||
Db     323 GPDDTNSEYESGSESEPDLSEDADSPWLLSNLVSRMISEGSSPIRCPGQCLSPAPRPPGE 382

Qy     326 --HPSISEEEEGFDCLSSPERAEPPGGGW---RGSLGEPPPPPRAS----------L 367
             |:    ::  |    || |  |||       :|  ||||  |:           |
Db     383 PVSPAGGAAQDSQD----PEAAAGPGGVELVDMETLCAPPPPAPAAPRPGPAQPGPCLFL 438

Qy     368 SSDT---------------------------------------------------------- 371
           |: |
Db     439 SNPTRDTITPLWAAPGRAARPGRACSAACSEEEDEEDDEEEEDAEDSAGSPGGRGTGPSA 498

Qy     372 ---SALSYDSVKYTLVVDEHAQLELVSLRPCFG-DYSDESDSATVYDNCASVSSPYESAI 427
              ::| ||:|||||||||| |||||||| | |  :  | ||
Db     499 PRDASLVYDAVKYTLVVDEHTQLELVSLRRCAGLGHDSEEDSG---------------- 541

Qy     428 GEEYEEAPRPQPPACL------SEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLFS 481
           ||  ||    :  | |       | |::|| ||: ||||||||::  ||||| |||||||
Db     542 GEASEE----EAGAALLGGGQVSGDTSPDSPDLTFSKKFLNVFVNSTSRSSSTESFGLFS 597

Qy     482 CIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFPA 541
           |::||||:||||||:|||:||| |||||:||||:||| : :|:|:  :||||| ||||||
```

```
Db    598 CLVNGEEREQTHRAVFRFIPRHPDELELDVDDPVLVEAEEDDFWFRGFNMRTGERGVFPA 657

Qy    542 YYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRLTV 601
          :|| |    : :    ::  ||::| |:||||:|| |:|| :|||||||||| |:|||
Db    658 FYAHAVPGPAKDLLGSKRSPCWVERFDVQFLGSVEVPCHQGNGILCAAMQKIATARKLTV 717

Qy    602 HFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGFIT 661
          |  ||:|| ||||:|||| :     :     :||||||:|||||| ||:|: ||||||
Db    718 HLRPPASCDLEISLRGVKLSLSGGGPE---FQRCSHFFQMKNISFCGCHPRNSCYFGFIT 774

Qy    662 KHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          |||   |||||||||::| : :|:|||||| ::|:: : | |||||||||
Db    775 KHPLLSRFACHVFVSQESMRPVAQSVGRAFLEYYQEHLAYACPTEDIYLE 824


RESULT 6
SP51_DROME
ID   SP51_DROME     STANDARD;      PRT;   490 AA.
AC   Q9W0K0; Q9NH69;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Eye developmental protein SP512.
GN   SP512 OR CG1200.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Serano T.L., Pendleton J.D., Rubin G.M.;
RT   "A reverse genetic screen for genes involved in Drosophila eye
RT   development.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- TISSUE SPECIFICITY: Expressed in the embryonic CNS and PNS and
CC       posterior to the morphogenetic furrow in the eye imaginal disk.
CC   -!- SIMILARITY: BELONGS TO THE JIP SCAFFOLD PROTEINS FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PID DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF231037; AAF34806.1; -.
DR   EMBL; AE003470; AAF47446.1; -.
DR   FlyBase; FBgn0040281; SP512.
DR   InterPro; IPR000050; PID_domain.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00018; SH3; 2.
DR   Pfam; PF00640; PID; 2.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00462; PTB; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS01179; PID; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   SH3 domain; Developmental protein.
FT   DOMAIN      150    153       POLY-ARG.
FT   DOMAIN      185    189       POLY-GLY.
FT   DOMAIN      251    260       PRO-RICH.
FT   DOMAIN      271    333       SH3.
FT   DOMAIN      344    479       PID.
FT   CONFLICT    164    171       SSILHLLG -> C (IN REF. 2).
SQ   SEQUENCE   490 AA;  53815 MW;  2744E85790EE043F CRC64;

  Query Match             14.2%;  Score 537.5;  DB 1;  Length 490;
  Best Local Similarity   39.4%;  Pred. No. 1.3e-20;
  Matches  114;  Conservative   60;  Mismatches   86;  Indels   29;  Gaps    8;

Qy    443 LSEDSTPDEP------DV---HFSKKFLNVFMSGRSRSSSAESFGLFSCIING------E 487
          : |||:||        ||   | :    | | |   : :|  |   |     |      |
Db    211 MDEDSSPDSERMQSLGDVDSGHSTAHSPNDFKSMSPQITSPVSQSPFPPPPFGGVPFGQLE 270

Qy    488 EQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFPAYYAIEV 547
            | ||| : :||||| ||:|||: | : |: :||| | |  |:||| :|:||: ||::
Db    271 MLEATHRGLHKFVPRHHDEIELEIGDAIYVQKEAEDLWCEGVNLRTGRQGIFPSAYAVDL 330

Qy    548 T-KEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRLTVHFNPP 606
          | :    | |     ::: : :||||:   |||  |:| |::||     :  : | |
Db    331 DYNEFDPTVQLVKK----ERYLLGYLGSVETLAHKGTGVVCQAVRKI-----VGEYGNSP 381

Qy    607 S--SCVLEISVRGVKIGVKADDSQEAKGNK--CSHFFQLKNISFCGYHPKNNKYFGFITK 662
          :  :|:||:| :|::: :: :|  |  |    :|: |||:||| :||:::: |||||
Db    382 TGQTCILEVSDQGLRMVDRSGPNQNKKDKKPCIDYFYSLKNVSFCAFHPRDHRFIGFITK 441

Qy    663 HPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          ||   |||||||   :||: :||:||||||:||::|:|   | ||||:|
Db    442 HPTVQRFACHVFKGSESTRPVAEAVGRAFQRFYQKFIETAYPIEDIYIE 490


RESULT 7
SYN1_HUMAN
ID   SYN1_HUMAN     STANDARD;      PRT;   705 AA.
AC   P17600; O75825;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Synapsin I (Brain protein 4.1).
```

```
GN   SYN1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=90243651; PubMed=2110562;
RA   Suedhof T.C.;
RT   "The structure of the human synapsin I gene and protein.";
RL   J. Biol. Chem. 265:7849-7852(1990).
RN   [2]
RP   SEQUENCE OF 1-125 FROM N.A.
RX   MEDLINE=90368667; PubMed=2118519;
RA   Sauerwald A., Hoesche C., Oschwald R., Kilimann M.W.;
RT   "The 5'-flanking region of the synapsin I gene. A G+C-rich, TATA- and
RT   CAAT-less, phylogenetically conserved sequence with cell
RT   type-specific promoter function.";
RL   J. Biol. Chem. 265:14932-14937(1990).
RN   [3]
RP   SEQUENCE OF 1-258 FROM N.A.
RA   Grafham D.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: NEURONAL PHOSPHOPROTEIN THAT COATS SYNAPTIC VESICLES,
CC       BINDS TO THE CYTOSKELETON, AND IS BELIEVED TO FUNCTION IN THE
CC       REGULATION OF NEUROTRANSMITTER RELEASE.
CC   -!- SUBCELLULAR LOCATION: SYNAPSE.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SYNAPSIN IA (SHOWN HERE) AND
CC       SYNAPSIN IB; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: SUBSTRATE OF AT LEAST FOUR DIFFERENT PROTEIN KINASES. IT IS
CC       PROBABLE THAT PHOSPHORYLATION PLAYS A ROLE IN THE REGULATION OF
CC       SYNAPSIN I IN THE NERVE TERMINAL.
CC   -!- SIMILARITY: BELONGS TO THE SYNAPSIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M58378; AAC41930.1; -.
DR   EMBL; M58321; AAC41930.1; JOINED.
DR   EMBL; M58341; AAC41930.1; JOINED.
DR   EMBL; M58351; AAC41930.1; JOINED.
DR   EMBL; M58353; AAC41930.1; JOINED.
DR   EMBL; M58359; AAC41930.1; JOINED.
DR   EMBL; M58371; AAC41930.1; JOINED.
DR   EMBL; M58372; AAC41930.1; JOINED.
DR   EMBL; M58373; AAC41930.1; JOINED.
DR   EMBL; M58374; AAC41930.1; JOINED.
DR   EMBL; M58375; AAC41930.1; JOINED.
DR   EMBL; M58376; AAC41930.1; JOINED.
DR   EMBL; M58377; AAC41930.1; JOINED.
DR   EMBL; M58378; AAC41931.1; ALT_SEQ.
DR   EMBL; M58321; AAC41931.1; JOINED.
DR   EMBL; M58341; AAC41931.1; JOINED.
DR   EMBL; M58351; AAC41931.1; JOINED.
DR   EMBL; M58353; AAC41931.1; JOINED.
DR   EMBL; M58359; AAC41931.1; JOINED.
DR   EMBL; M58371; AAC41931.1; JOINED.
DR   EMBL; M58372; AAC41931.1; JOINED.
DR   EMBL; M58373; AAC41931.1; JOINED.
DR   EMBL; M58374; AAC41931.1; JOINED.
DR   EMBL; M58375; AAC41931.1; JOINED.
DR   EMBL; M58376; AAC41931.1; JOINED.
DR   EMBL; M58377; AAC41931.1; JOINED.
DR   EMBL; M55301; AAA60608.1; -.
DR   EMBL; AL009172; CAA15657.1; -.
DR   PIR; A35363; A35363.
DR   PIR; B35363; B35363.
DR   HSSP; P17599; 1AUX.
DR   Genew; HGNC:11494; SYN1.
DR   MIM; 313440; -.
DR   InterPro; IPR001359; Synapsin.
DR   Pfam; PF02078; Synapsin; 1.
DR   Pfam; PF02750; Synapsin_C; 1.
DR   PRINTS; PR01368; SYNAPSIN.
DR   PROSITE; PS00415; SYNAPSIN_1; 1.
DR   PROSITE; PS00416; SYNAPSIN_2; 1.
KW   Synapse; Phosphorylation; Neurone; Repeat; Actin-binding;
KW   Alternative splicing.
FT   DOMAIN        1     28       A.
FT   DOMAIN       29    112       B (LINKER).
FT   DOMAIN      113    420       C (ACTIN-BINDING AND SYNAPTIC-VESICLE
FT                                BINDING).
FT   DOMAIN      421    655       D (PRO-RICH LINKER).
FT   DOMAIN      656    705       E.
FT   MOD_RES       9      9       PHOSPHORYLATION (BY PKA AND CAMK1).
FT   MOD_RES     568    568       PHOSPHORYLATION (BY CAMK2).
FT   MOD_RES     605    605       PHOSPHORYLATION (BY CAMK2).
FT   VARSPLIC    661    669       NKSQSLTNA -> KASPAQAQP (IN SYNAPSIN IB).
FT   VARSPLIC    670    705       MISSING (IN SYNAPSIN IB).
FT   CONFLICT    138    138       G -> E (IN REF. 3).
SQ   SEQUENCE   705 AA;  73954 MW;  487831123FF6882F CRC64;

  Query Match             5.2%;  Score 195.5;  DB 1;  Length 705;
  Best Local Similarity   25.8%;  Pred. No. 0.0033;
  Matches   73;  Conservative   25;  Mismatches  128;  Indels   57;  Gaps   10;

Qy    100 TGDTPGAEDDEEDDDEERAARRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPK 159
          |   |     ::    :    :|| || :     :     |||      || :|    |:
Db    448 TSQQPAGPPAQQRPPPQGGPPQPGPGPQRQGPPLQQRPPPQGQQHLSGLGPPAGSPL-PQ 506

Qy    160 RPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPP 219
          |  :    || | ||         |          | | |: |   ||      |
Db    507 RLPSPTSAPQQPASQAAPPTQGQG----------RQSRPVAGGPGAPPAARPPASPSPQR 556

Qy    220 QSGPAPTTDRGTSTDSPCRRSTATQMAPPG-----GPPAAPPGGRGHSHRDRIHYQADVR 274
          |:|| |   | ||    |     :   ||||     |||  |||  | :
Db    557 QAGP-PQATRQTSVSGPAPPKASG--APPGGQQRQGPPQKPPGPAGPTR---------- 602

Qy    275 LEATEEIYLTPVQR--PPDAAEPT-----SAFLPPTESRMSVSSDPDPAAYPSTAGRPHP 327
           :|::     || |  ||   :|       |  |  :     | |  | |  :  | |||
Db    603 -QASQ---AGPVPRTGPPTTQQPRPSGPGPAGAPKPQLAQKPSQDVPPPATAAAGGPPHP 658

Qy    328 SISEEEEGFDCLSSPERAEPPGGGWRGSLGEPPPPPRASLSSD 370
           ::: :   :  : ||               | ||| ||| |
Db    659 QLNKSQSLTNAFNLPE---------------PAPPRPSLSQD 685


RESULT 8
BAT2_HUMAN
ID   BAT2_HUMAN     STANDARD;      PRT;  2142 AA.
AC   P48634;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Large proline-rich protein BAT2 (HLA-B-associated transcript 2).
GN   BAT2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=T-cell;
RX   MEDLINE=90192810; PubMed=2156268;
RA   Banerji J., Sands J., Strominger J.L., Spies T.;
RT   "A gene pair from the human major histocompatibility complex encodes
RT   large proline-rich proteins with multiple repeated motifs and a
RT   single ubiquitin-like domain.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:2374-2378(1990).
RN   [2]
```

```
RP   SEQUENCE OF 1-1860 FROM N.A.
RX   MEDLINE=93272029; PubMed=8499947;
RA   Iris F.J.M., Bougueleret L., Prieur S., Caterina D., Primas G.,
RA   Perrot V., Jurka J., Rodriguez-Tome P., Claverie J.-M., Dausset J.,
RA   Cohen D.;
RT   "Dense Alu clustering and a potential new member of the NF kappa B
RT   family within a 90 kilobase HLA class III segment.";
RL   Nat. Genet. 3:137-145(1993).
CC   -!- FUNCTION: UNKNOWN.
CC   -!- TISSUE SPECIFICITY: LIMITED TO CELL-LINES OF LEUKEMIC ORIGIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33509; AAA35585.1; -.
DR   EMBL; M33518; AAA35586.1; -.
DR   EMBL; M33512; AAA35586.1; JOINED.
DR   EMBL; Z15025; CAA78744.1; -.
DR   PIR; B35098; B35098.
DR   PIR; S36152; S36152.
DR   Genew; HGNC:13918; BAT2.
DR   MIM; 142580; -.
KW   Repeat.
FT   DOMAIN      519    524       POLY-PRO.
FT   DOMAIN      636    657       GLN-RICH.
FT   DOMAIN      684    688       POLY-PRO.
FT   DOMAIN      699    704       POLY-PRO.
FT   DOMAIN      814    821       POLY-PRO.
FT   DOMAIN     1340   1345       POLY-GLY.
FT   DOMAIN     1398   1403       POLY-GLY.
FT   DOMAIN     1436   1442       POLY-PRO.
FT   DOMAIN     1982   1991       POLY-PRO.
FT   DOMAIN       41   1795       4 X 57 AA TYPE A REPEATS.
FT   REPEAT       41     95       1-1.
FT   REPEAT       98    154       1-2.
FT   REPEAT      281    337       1-3.
FT   REPEAT     1740   1795       1-4.
FT   DOMAIN      337    549       2 X TYPE B REPEATS.
FT   REPEAT      337    418       2-1.
FT   REPEAT      476    549       2-2.
FT   DOMAIN     1899   2089       3 X 50 AA TYPE C REPEATS.
FT   REPEAT     1899   1948       3-1.
FT   REPEAT     1965   2014       3-2.
FT   REPEAT     2040   2089       3-3.
FT   CONFLICT     57     57       R -> A (IN REF. 2).
FT   CONFLICT    109    109       Q -> S (IN REF. 2).
FT   CONFLICT    414    414       P -> PPHRGPAGNWGPP (IN REF. 2).
FT   CONFLICT    532    532       T -> K (IN REF. 2).
FT   CONFLICT    682    682       Q -> K (IN REF. 2).
FT   CONFLICT    730    730       E -> D (IN REF. 2).
FT   CONFLICT    750    750       L -> R (IN REF. 2).
FT   CONFLICT    834    834       A -> T (IN REF. 2).
FT   CONFLICT   1035   1035       G -> A (IN REF. 2).
FT   CONFLICT   1068   1068       M -> L (IN REF. 2).
FT   CONFLICT   1285   1285       P -> R (IN REF. 2).
FT   CONFLICT   1400   1400       G -> A (IN REF. 2).
FT   CONFLICT   1611   1611       T -> S (IN REF. 2).
FT   CONFLICT   1729   1729       G -> A (IN REF. 2).
SQ   SEQUENCE   2142 AA;  227840 MW;  32DDF16B9B52420A CRC64;

  Query Match             5.1%;  Score 194.5;  DB 1;  Length 2142;
  Best Local Similarity   23.2%;  Pred. No. 0.012;
  Matches  118;  Conservative   33;  Mismatches  170;  Indels  187;  Gaps   22;

Qy      13 AASPPAASPFLGLHIASPPNFRLTHDISLEEFED-EDLSEITDECGISLQCKDTLSLRPP 71
           | :|| |:        ||       | : ||:|   : |: | |     :    : | |
Db    1281 APAPPRAA------AKSPDLSNQNSDQANEEWETASESSDFTSERRGDKEAPPPVLLTPK 1334

Qy      72 RAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAARRPGAGPPKAES 131
             |    |||||   : | | : ||     |        :    ||  |||| |     |
Db    1335 AVGTPGGGGGGAVPGISA-MSRGDLSQRAKDLSKRSFSSQRPGMERQNRRPGPGGKAGSS 1393

Qy     132 GQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDR 191
           |   :| | |   |   ||| ||                                  | ||
Db    1394 GS--SSGGGGGGPGGRTGPGRGD----------------------------KRSWPSP 1421

Qy     192 VSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRRSTA--------- 242
            :||  |    |: ||         |||   | |::      |     : |
Db    1422 KNRSRPP----EERPP------GLPLPP---PPPSSSAVFRLDQVIHSNPAGIQQALAQL 1468

Qy     243 -----TQMAPPG------GPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPD 291
                :  || |      ||| || |                                | |||
Db    1469 SSRQGSVTAPGGHPRHKPGPPQAPQG----------------------------PSPRPPT 1501

Qy     292 AAEPTSAFLPPTESRMSVSSDP----------------DPAAY----PSTAGRP--HPS 328
            ||             :||||                | |      |    ||   |
Db    1502 RYEPQRV-------NSGLSSDPHFEEPGPMVRGVGGTPRDSAGVSPFPPKRRERPPRKPE 1554

Qy     329 ISEEEEGFDCLSSPERAEPPGGGWRGSLGEPPPPPRASLSSDTSALS---YDSVKYTLVV 385
           : :||      | |    |   |: || | | |   | : | ||:   :: :
Db    1555 LLQEE------SLP---PPHSSGFLGSKPEGPGPQAESRDTGTEALTPHIWNRL------ 1599

Qy     386 DEHAQLELVSLRPCFGDYSDESDSATVYDNCASVSSPYESAIGEEYEEAPRPQPPACLSE 445
           |      | ||           |  :      ||:|   | |  ::
Db    1600 --HTATSRKSYRP----------TSMEPWMEPLSPFEDVAGTEMSQS---------- 1634

Qy     446 DSTPDEPDVHFSKKFLNVFMSGRSRSSS 473
           ||  |              :|| |: ||
Db    1635 DSGVD-------------LSGDSQVSS 1648


RESULT 9
WAIP_HUMAN
ID   WAIP_HUMAN     STANDARD;      PRT;   503 AA.
AC   O43516; Q9UNP1; Q15220;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Wiskott-Aldrich syndrome protein interacting protein (WASP interacting
DE   protein) (PRPL-2 protein).
GN   WASPIP OR WIP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98070810; PubMed=9405671;
RA   Ramesh N., Anton I.M., Hartwig J.H., Geha R.S.;
RT   "WIP, a protein associated with Wiskott-Aldrich syndrome protein,
RT   induces actin polymerization and redistribution in lymphoid cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:14671-14676(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Tonsil;
RA   Kreideweiss S., Delany-Heiken P., Nordheim A., Ruhlmann A.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 192-503 FROM N.A., AND VARIANT ALA-495.
RX   MEDLINE=99218549; PubMed=10202051;
RA   Stewart D.M., Tian L., Nelson D.L.;
RT   "Mutations that cause the Wiskott-Aldrich syndrome impair the
RT   interaction of Wiskott-Aldrich syndrome protein (WASP) with WASP
RT   interacting protein.";
RL   J. Immunol. 162:5019-5024(1999).
CC   -!- FUNCTION: MAY HAVE DIRECT ACTIVITY ON THE ACTIN CYTOSKELETON.
CC       INDUCES ACTIN POLYMERIZATION AND REDISTRIBUTION.
CC   -!- SUBUNIT: BINDS TO WASP, PROFILIN AND ACTIN.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PERIPHERAL BLOOD
CC       MONONUCLEAR CELLS, SPLEEN, PLACENTA, SMALL INTESTIN, COLON,
CC       THYMUS. LOWER EXPRESSION IN OVARY, HEART, BRAIN, LUNG, LIVER,
```

```
CC       SKELETAL MUSCLE, KIDNEY, PANCREAS, PROSTATE, TESTIS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF031588; AAC03767.1; -.
DR   EMBL; AF106062; AAD45972.1; -.
DR   EMBL; X86019; CAA60014.1; -.
DR   Genew; HGNC:12736; WASPIP.
DR   MIM; 602357; -.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF02205; WH2; 1.
DR   SMART; SM00246; WH2; 1.
KW   Actin-binding; Repeat.
FT   DOMAIN        2     13       POLY-PRO.
FT   DOMAIN       64     96       GLY-RICH.
FT   DOMAIN      241    244       POLY-SER.
FT   DOMAIN      264    433       PRO-RICH.
FT   REPEAT      352    361       XRSGPXPPXP MOTIF 1.
FT   REPEAT      374    383       XRSGPXPPXP MOTIF 2.
FT   REPEAT      410    419       XRSGPXPPXP MOTIF 3.
FT   SITE         45     48       ACTIN BINDING.
FT   VARIANT     495    495       G -> A.
FT                                /FTId=VAR_010295.
FT   CONFLICT    303    309       PHRPHLR -> SSQAPPP (IN REF. 3).
FT   CONFLICT    360    360       P -> PV (IN REF. 2).
FT   CONFLICT    487    503       GSNRRERGGPPLPPIPR -> EYFCQGF (IN REF. 2).
SQ   SEQUENCE   503 AA;  51489 MW;  43EB88674DD3BF1A CRC64;

  Query Match             5.1%;  Score 193.5;  DB 1;  Length 503;
  Best Local Similarity   21.7%;  Pred. No. 0.0029;
  Matches  124;  Conservative   39;  Mismatches  207;  Indels  201;  Gaps   23;

Qy      16 PPAASPFLGLHIASPPNFRL--THDISLEEFEDEDLSEITDECGISLQCKDTLSL----- 68
           |||  |        || | |  |   :| : |    : :  :     : | |::
Db       6 PPAPPP--------PPTFALANTEKPTLNKTEQAGRNALLSDISKGKKLKKTVTNDRSAP 57

Qy      69 ---RPPRAGLLSAGGG-------------------------GAGSRLQAEMLQMDLIDAT 100
             :|  ||    |||                          | |   || |
Db      58 ILDKPKGAGAGGGGGGFGGGGGFGGGGGGGGGGSFGGGGPPGLGGLFQAGM--------- 108

Qy     101 GDTPGAEDDEEDDDEERAARRPGAGPPKAESGQEPASRGQG------QSQGQSQGPGSGD 154
           |        |::   :| |   |     : :| | |     |       | |   ||
Db     109 ---PKLRSTANRDNDSGGSRPPLLPPGGRSTSAKPFSPPSGPGRFPVPSPGHRSGPPEPQ 165

Qy     155 TYR--PKRPTTLNL-------FPQVPRS-QDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQ 204
             | | ||   :                | ||  | :|:|                | | |:  |
Db     166 RNRMPPPRPDVGSKPDSIPPPVPSTPRPIQSSLHN--------------RGSPPVPGG-- 209

Qy     205 TPPHEHICLSEELPPQSGPAPTT-----DRGTS------TDSPCRRSTATQMAPP----- 248
                 | |  | ||      :|||:         ||    |:     ||
Db     210 ------------PRQPSPGPTPPPFPGNRGTALGGGSIRQSPLSSSSPFSNRPPLPPTP 256

Qy     249 -------GGPPAAPPGGRGHSHRDRI------HYQADVRLEATEEIYLTPVQRPPDAAEP 295
                  ||  | | |   ||: :      : :  |           |  |||  : |
Db     257 SRALDDKPPPPPPPVGNRPSIHREAVPPPPPQNNKPPVPSTPRPSAPHRPHLRPPPPSRP 316

Qy     296 TSAFLPPTES----------RMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPER 344
             |||: |             :|:||   |   |  :|   |  ||          |
Db     317 GPPPLPPSSSGNDETPRLPQRNLSLSSSTPPLPSPGRSGPLPPPPSER-------PPPPV 369

Qy     345 AEPPGGGWRGSLGEPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYS 404
           :|||    | |  |||||:   | : ||                       |
Db     370 RDPPGRS--GPL--PPPPPVSRNGSTSRAL----------------------PATPQLP 402

Qy     405 DESDSATVYDNCASVSSPYESAIGEEYEEAPRPQPPA----------CLSE-------- 445
             |   :          |   : |     || | ||:          |  |
Db     403 SRSGVDSPRSGPRPPLPPDRPSAG-----APPPPPPSTSIRNGFQDSPCEDEWESRFYFH 457

Qy     446 ---DSTPDEPDVHFSKKFLNVFMSGRSRSSS 473
                 |  | || |  :| : :      ||| |
Db     458 PISDLPPPEPYVQTTKSYPSKLARNESRSGS 488


RESULT 10
SYN1_BOVIN
ID   SYN1_BOVIN     STANDARD;      PRT;   706 AA.
AC   P17599;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Synapsin I.
GN   SYN1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=89388265; PubMed=2506642;
RA   Suedhof T.C., Czernik A.J., Kao H.-T., Takei K., Johnston P.A.,
RA   Horiuchi A., Kanazir S.D., Wagner M.A., Perin M.S., de Camilli P.,
RA   Greengard P.;
RT   "Synapsins: mosaics of shared and individual domains in a family of
RT   synaptic vesicle phosphoproteins.";
RL   Science 245:1474-1480(1989).
RN   [2]
RP   PHOSPHORYLATION SITE BY PROLINE-DIRECTED KINASE.
RX   MEDLINE=90216728; PubMed=2108963;
RA   Hall F.L., Mitchell J.P., Vulliet P.R.;
RT   "Phosphorylation of synapsin I at a novel site by proline-directed
RT   protein kinase.";
RL   J. Biol. Chem. 265:6944-6948(1990).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 112-417.
RX   MEDLINE=98130589; PubMed=9463376;
RA   Esser L., Wang C.R., Hosaka M., Smagula C.S., Suedhof T.C.,
RA   Deisenhofer J.;
RT   "Synapsin I is structurally similar to ATP-utilizing enzymes.";
RL   EMBO J. 17:977-984(1998).
CC   -!- FUNCTION: NEURONAL PHOSPHOPROTEIN THAT COATS SYNAPTIC VESICLES,
CC       BINDS TO THE CYTOSKELETON, AND IS BELIEVED TO FUNCTION IN THE
CC       REGULATION OF NEUROTRANSMITTER RELEASE.
CC   -!- SUBCELLULAR LOCATION: SYNAPSE.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SYNAPSIN IA (SHOWN HERE) AND
CC       SYNAPSIN IB; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: SUBSTRATE OF AT LEAST FOUR DIFFERENT PROTEIN KINASES. IT IS
CC       PROBABLE THAT PHOSPHORYLATION PLAYS A ROLE IN THE REGULATION OF
CC       SYNAPSIN I IN THE NERVE TERMINAL.
CC   -!- SIMILARITY: BELONGS TO THE SYNAPSIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M27810; AAA30761.1; -.
DR   EMBL; M27811; AAA30762.1; -.
DR   PIR; A35758; A35758.
DR   PIR; E30411; E30411.
DR   PIR; F30411; F30411.
DR   PDB; 1AUV; 18-MAR-98.
DR   PDB; 1AUX; 18-MAR-98.
DR   GlycoSuiteDB; P17599; -.
DR   InterPro; IPR001359; Synapsin.
DR   Pfam; PF02078; Synapsin; 1.
```

```
DR   Pfam; PF02750; Synapsin_C; 1.
DR   PRINTS; PR01368; SYNAPSIN.
DR   PROSITE; PS00415; SYNAPSIN_1; 1.
DR   PROSITE; PS00416; SYNAPSIN_2; 1.
KW   Synapse; Phosphorylation; Neurone; Repeat; Actin-binding;
KW   Alternative splicing; 3D-structure.
FT   DOMAIN        1     28       A.
FT   DOMAIN       29    112       B (LINKER).
FT   DOMAIN      113    420       C (ACTIN-BINDING AND SYNAPTIC-VESICLE
FT                                BINDING).
FT   DOMAIN      421    656       D (PRO-RICH LINKER).
FT   DOMAIN      657    706       E.
FT   MOD_RES       9      9       PHOSPHORYLATION (BY PKA AND CAMK1).
FT   MOD_RES     551    551       PHOSPHORYLATION (BY PDPK).
FT   MOD_RES     568    568       PHOSPHORYLATION (BY CAMK2).
FT   MOD_RES     605    605       PHOSPHORYLATION (BY CAMK2).
FT   VARSPLIC    662    670       NKSQSLTNA -> KASPAQAQP (IN SYNAPSIN IB).
FT   VARSPLIC    671    706       MISSING (IN SYNAPSIN IB).
SQ   SEQUENCE   706 AA;  74518 MW;  89373750BF014340 CRC64;

  Query Match             4.9%;  Score 184.5;  DB 1;  Length 706;
  Best Local Similarity   27.3%;  Pred. No. 0.012;
  Matches   74;  Conservative   23;  Mismatches  105;  Indels   69;  Gaps   12;

Qy     122 PGAGPPKAESG---------QEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPR 172
           |  |||:   |           | |  :||    |     ||    |   ||::   || |
Db     463 PQGGPPQPGPGPQRQGPPLQQRPTPQGQQHLSGLGPPAGSPLPQRLPSPTSV---PQQPA 519

Qy     173 SQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTS 232
           || |      |          | | |:  |   ||      |     |:|| |   | ||
Db     520 SQATPMTQGQG----------RQSRPVAGGPGAPPATRPPASPSPQRQAGP-PQATRQTS 568

Qy     233 T--DSPCRRSTATQMAPPG-----GPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTP 285
             :| : |      |||     |||  |||  | : :               :    |
Db     569 VSGQAPPKASG----VPPGGQQRQGPPQKPPGPAGPTRQ------------ASQAGPMP 611

Qy     286 VQRPPDAAEP-TSAFLP---PTESRMS--VSSDPDPAAYPSTAGRPHPSISEEEEGFDCL 339
             ||   :|  |   |   ||: :::   | |  | |  :  | ||| ::: :   :
Db     612 RTGPPTTQQPRPSGPGPAGRPTKPQLAQKPSQDVPPPATAAAGGPPHPQLNKSQSLTNAF 671

Qy     340 SSPERAEPPGGGWRGSLGEPPPPPRASLSSD 370
           : ||                | ||| ||| |
Db     672 NLPE---------------PAPPRPSLSQD 686


RESULT 11
TENS_CHICK
ID   TENS_CHICK     STANDARD;      PRT;  1744 AA.
AC   Q04205; Q91007; Q92011;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Tensin.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=94350987; PubMed=8071358;
RA   Lo S.H., An Q., Bao S., Wong W.K., Liu Y., Janmey P.A., Hartwig J.H.,
RA   Chen L.B.;
RT   "Molecular cloning of chick cardiac muscle tensin. Full-length cDNA
RT   sequence, expression, and characterization.";
RL   J. Biol. Chem. 269:22310-22319(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=95204530; PubMed=7896874;
RA   Chuang J.Z., Lin D.C., Lin S.;
RT   "Molecular cloning, expression, and mapping of the high affinity
RT   actin-capping domain of chicken cardiac tensin.";
RL   J. Cell Biol. 128:1095-1109(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Chen L.B.;
RL   Submitted (XXX-1991) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1469-1744 FROM N.A.
RC   TISSUE=Embryonic chondrocytes, and Embryonic heart;
RX   MEDLINE=94039118; PubMed=8223621;
RA   van de Werken R., Gennari M., Tavella S., Bet P., Molina F.,
RA   Lin S., Cancedda R., Castagnola P.;
RT   "Modulation of tensin and vimentin expression in chick embryo
RT   developing cartilage and cultured differentiating chondrocytes.";
RL   Eur. J. Biochem. 217:781-790(1993).
RN   [5]
RP   SH2 DOMAIN.
RX   MEDLINE=91220073; PubMed=1708917;
RA   Davis S., Lu M.L., Lo S.H., Lin S., Butler J.A., Druker B.J.,
RA   Roberts T.M., An Q., Chen L.B.;
RT   "Presence of an SH2 domain in the actin-binding protein tensin.";
RL   Science 252:712-715(1991).
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL MIGRATION, CARTILAGE DEVELOPMENT
CC       AND IN LINKING SIGNAL TRANSDUCTION PATHWAYS TO THE CYTOSKELETON.
CC       BINDS TO ACTINS AND PHOSPHORYLATED PROTEINS IN SRC-TRANSFORMED
CC       CELLS. MAY BIND ACTIN WITH CAPPING AND BUNDLING PROPERTIES.
CC   -!- SUBCELLULAR LOCATION: LOCALIZES TO ADHERENS JUNCTIONS.
CC   -!- TISSUE SPECIFICITY: HEART, GIZZARD, LUNG AND SKELETAL MUSCLE.
CC   -!- PTM: TYROSINE-PHOSPHORYLATED.
CC   -!- SIMILARITY: CONTAINS 1 TENSIN DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M96625; AAA59053.1; -.
DR   EMBL; L06662; AAA73949.1; ALT_INIT.
DR   EMBL; Z18529; CAA79215.1; ALT_INIT.
DR   EMBL; M74165; AAA49087.1; -.
DR   EMBL; X66286; CAA46992.1; -.
DR   HSSP; P16277; 1BLK.
DR   InterPro; IPR000340; DS_phosphatase.
DR   InterPro; IPR000050; PID_domain.
DR   InterPro; IPR000980; SH2.
DR   Pfam; PF00017; SH2; 1.
DR   ProDom; PD000093; SH2; 1.
DR   SMART; SM00462; PTB; 1.
DR   SMART; SM00012; PTPc_DSPc; 1.
DR   SMART; SM00252; SH2; 1.
DR   PROSITE; PS50001; SH2; 1.
KW   Actin-binding; Cytoskeleton; SH2 domain; Phosphorylation.
FT   DOMAIN       66    342       TENSIN.
FT   DOMAIN     1472   1581       SH2.
FT   CONFLICT     49     49       C -> R (IN REF. 2).
FT   CONFLICT     61     61       M -> T (IN REF. 2).
FT   CONFLICT     88     88       Q -> PR (IN REF. 3).
FT   CONFLICT    404    404       A -> T (IN REF. 2).
FT   CONFLICT    452    452       T -> A (IN REF. 2).
FT   CONFLICT    508    509       DV -> EL (IN REF. 2).
FT   CONFLICT    522    522       A -> P (IN REF. 2).
FT   CONFLICT    664    664       R -> A (IN REF. 2).
FT   CONFLICT    666    666       A -> T (IN REF. 2).
FT   CONFLICT    875    875       P -> A (IN REF. 2).
FT   CONFLICT    909    909       P -> T (IN REF. 2).
FT   CONFLICT   1102   1113       MISSING (IN REF. 3).
FT   CONFLICT   1240   1240       G -> A (IN REF. 2).
FT   CONFLICT   1480   1480       E -> D (IN REF. 2).
FT   CONFLICT   1711   1711       D -> E (IN REF. 2).
```

```
SQ   SEQUENCE   1744 AA;  187214 MW;  5C3C8B6211935524 CRC64;

  Query Match             4.7%;  Score 179;  DB 1;  Length 1744;
  Best Local Similarity   19.8%;  Pred. No. 0.058;
  Matches  177;  Conservative   97;  Mismatches  312;  Indels  306;  Gaps   39;

Qy      13 AASPPAASPFLGLHIASPPN------FRLTHDISLEEFEDEDLSEITDECGISLQCKDTL 66
           |:  || |    | ::|||:       :| |   || :||           :|   :: |
Db     928 ASYSPAGSQ--QLLVSSPPSPTAPAQSQLPHK-GLESYED----------LSRSGEEPL 973

Qy      67 SLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAARRP---- 122
           :|     ||::    |  || ::   :   : |   ||      |:   |  : | |
Db     974 NLE----GLVAHRVAGVQSREKSP--EESTVPARRRTPSDSHYEKSSPEPGSPRSPTVLS 1027

Qy     123 ----------GAGPPK---------AESGQEPASRGQGQSQGQSQGPGSGDT------YR 157
                     | ||         |    | ||     | |    ||   |        :
Db    1028 PEVVSTIAANPGGRPKEPHLHSYKEAFEEMESASPSSLTSGGVRSPPGLAKTPLSALGLK 1087

Qy     158 PKRPTTLNLFP------------QVPRS-------------------------QDTLNN 179
           |  |  : | |             : |||                        :    |
Db    1088 PHNPADILLHPVGELEGEAGADSEEEPRSYVESVARTATTGRAGNLPAAQPVGLEVPARN 1147

Qy     180 NSLGKKHSWQDRVSRSSSPLKT---------------GEQTPPHEHICLSEELPPQSGP 223
           : |   :    || :|||: :                 |  ||||  :  :    |
Db    1148 GAFGNSFTVPSPVS-TSSPIHSVDGASLRSYPSEGSPHGTVTPPHA-VAETAYRSPMVSQ 1205

Qy     224 APTTDRGTSTDSP---------------------------------------------- 236
           |:      | ||
Db    1206 TPSAHSSYQTSSPSSFQAGTLGSPYASPDYPDGRGGFQPDPQARQQPQVSVVGVHALPGS 1265

Qy     237 ---CRRSTATQMAP----------------PGGP--------PAAPPGGRGHSHRDRIHY 269
              |: ||   |                 || |        | ||||    :     |
Db    1266 PRTLHRTVATNTPPSPGFGRRAANPAVASVPGSPGLGRHTVSPHAPPGSPSLAR----HQ 1321

Qy     270 QADVRLEATEEIYLTPVQRPPDAAEPTSA--FLPPTESRMSV-----------SSDPDP 315
           | |   :    | :| :| | :  :||  : ||:     |             || ||
Db    1322 MAAVPPGSPMYGYSSPEERRPTLSRQSSASGYQPPSTPSFPVSPAYYPGTSTPHSSSPDS 1381

Qy     316 AAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSL-------GEPPPPPRASLS 368
           |||   :  | |:: |:      :|: ||:          ||       |:   |  : :|
Db    1382 AAYRQGSPTPQPALPEKRR----MSAGERS--------NSLPNYATVNGKASSPLSSGMS 1429

Qy     369 SDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASVSSP----YE 424
           | :|  |  :  :||             | |  :|    |     |   |       |:
Db    1430 SPSSG-SAVAFSHTL-------------PDFSKFSMPDISPETRANVKFVQDTSKYWYK 1474

Qy     425 SAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLFSCII 484
             |  |   |        | :|  |    :  |  |   :      :|        : :
Db    1475 PDISREQAIA--------LLKDREPGAFIIRDSHSFRGAYGLAMKVASPPP-----TVMQ 1521

Qy     485 NGEEQEQTHRAIFRFV----PRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFP 540
           :: : |:  :  |:    ||  :|:   ::|    | |  | :    :    : | |
Db    1522 QNKKGDITNELVRHFLIETSPRGV-KLKGCPNEPNFGCLSALVYQHSIMPLALPCKLVIP 1580

Qy     541 AYYAIEVTKEPEHMAALAKNS--DWVDQ---FRVKFLGSVQVPYHKGNDVLCAAMQKI-- 593
             :|  |:    ||  ||  | : |     | |: ||::    |   :  |: :
Db    1581 DRDPMEEKKD----AASTTNSATDLLKQGAACNVLFINSVEMESLTGPQAISKAVAETLV 1636

Qy     594 --ATTRRLTVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQ----LKNISFC 647
               |     |||        ::| :|: :             |:   ||:    |  ::||
Db    1637 ADPTPTATIVHF--------KVSAQGITL----------TDNQRKLFFRRHYPLNTVTFC 1678

Qy     648 GYHPKNNKY----------FGFIT-KHPADHRFACHVFVSEDSTKALAESV 687
             |:  |:           |||:  |  :     ||:|   |  :  |  |
Db    1679 DLDPQERKWTKTDGSGPAKLFGFVARKQGSTTDNVCHLFAELDPDQPAAAIV 1730


RESULT 12
IRS1_HUMAN
ID   IRS1_HUMAN     STANDARD;      PRT;  1242 AA.
AC   P35568;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Insulin receptor substrate-1 (IRS-1).
GN   IRS1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=93292738; PubMed=8513971;
RA   Araki E., Sun X.J., Haag B.L. III, Chuang L.M., Zhang Y.,
RA   Yang-Feng T.L., White M.F., Kahn C.R.;
RT   "Human skeletal muscle insulin receptor substrate-1. Characterization
RT   of the cDNA, gene, and chromosomal localization.";
RL   Diabetes 42:1041-1054(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92181456; PubMed=1311924;
RA   Nishiyama M., Wands J.R.;
RT   "Cloning and increased expression of an insulin receptor substrate-1-
RT   like gene in human hepatocellular carcinoma.";
RL   Biochem. Biophys. Res. Commun. 183:280-285(1992).
RN   [3]
RP   VARIANTS PRO-512 AND ARG-971.
RX   MEDLINE=93390176; PubMed=8104271;
RA   Almind K., Bjoerbaek C., Vestergaard H., Hansen T., Echwald S.,
RA   Pedersen O.;
RT   "Aminoacid polymorphisms of insulin receptor substrate-1 in
RT   non-insulin-dependent diabetes mellitus.";
RL   Lancet 342:828-832(1993).
RN   [4]
RP   VARIANT NIDDM GLY-723 DEL.
RX   MEDLINE=96303710; PubMed=8723689;
RA   Esposito D.L., Mammarella S., Ranieri A., della Loggia F., Capani F.,
RA   Consoli A., Mariani-Costantini R., Caramia F.G., Cama A., Battista P.;
RT   "Deletion of Gly723 in the insulin receptor substrate-1 of a patient
RT   with noninsulin-dependent diabetes mellitus.";
RL   Hum. Mutat. 7:364-366(1996).
RN   [5]
RP   VARIANTS NIDMM TYR-1043 AND TYR-1095.
RA   Mammarella S., Creati B., Esposito D.L., Arcuri P., della Loggia F.,
RA   Capani F., Mariani-Costantini R., Caramia F.G., Battista P., Cama A.;
RT   "Novel allele of the insulin receptor substrate-1 bearing two
RT   non-conservative amino acid substitutions in a patient with
RT   noninsulin-dependent diabetes mellitus.";
RL   Hum. Mutat. 11:411-411(1998).
RN   [6]
RP   STRUCTURE BY NMR OF 157-267.
RX   MEDLINE=96185451; PubMed=8599766;
RA   Zhou M.-M., Huang B., Olejniczak E.T., Meadows R.P., Shuker S.B.,
RA   Miyazaki M., Trueb T., Shoelson S.E., Fesik S.W.;
RT   "Structural basis for IL-4 receptor phosphopeptide recognition by the
RT   IRS-1 PTB domain.";
RL   Nat. Struct. Biol. 3:388-393(1996).
CC   -!- FUNCTION: MAY MEDIATE THE CONTROL OF VARIOUS CELLULAR PROCESSES
CC       BY INSULIN. WHEN PHOSPHORYLATED BY THE INSULIN RECEPTOR BINDS
CC       SPECIFICALLY TO VARIOUS CELLULAR PROTEINS CONTAINING SH2 DOMAINS
CC       SUCH AS PHOSPHATIDYLINOSITOL 3-KINASE P85 SUBUNIT OR GRB-2.
CC   -!- DISEASE: POLYMORPHISMS IN IRS1 MAY BE INVOLVED IN THE ETIOLOGY OF
CC       A SUBSET OF LATE-ONSET NON-INSULIN-DEPENDENT DIABETES MELLITUS
CC       (NIDDM).
CC   -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PTB DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; S85963; AAB21608.1; -.
DR   EMBL; S62539; AAB27175.1; -.
DR   PIR; JS0670; JS0670.
DR   PDB; 1IRS; 15-MAY-97.
DR   Genew; HGNC:6125; IRS1.
DR   MIM; 147545; -.
DR   InterPro; IPR002404; Insulin_receptS1.
DR   InterPro; IPR001849; PH.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF02174; IRS; 1.
DR   PRINTS; PR00628; INSULINRSI.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00310; PTBI; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
KW   Phosphorylation; Polymorphism; Diabetes mellitus; Disease mutation;
KW   3D-structure.
FT   DOMAIN       12    115       PH.
FT   DOMAIN      157    267       PTB.
FT   DOMAIN      128    134       POLY-GLY.
FT   DOMAIN      391    398       POLY-SER.
FT   DOMAIN      680    686       POLY-SER.
FT   DOMAIN      807    815       POLY-SER.
FT   DOMAIN      877    882       POLY-GLN.
FT   DOMAIN     1035   1038       POLY-SER.
FT   DOMAIN     1126   1130       POLY-GLY.
FT   DOMAIN     1131   1135       POLY-SER.
FT   DOMAIN     1197   1207       POLY-PRO.
FT   MOD_RES      99     99       PHOSPHORYLATION (BY CK2)
FT                                (BY SIMILARITY).
FT   MOD_RES     465    465       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   MOD_RES     612    612       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   MOD_RES     632    632       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   MOD_RES     896    896       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   MOD_RES     941    941       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   MOD_RES     989    989       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   MOD_RES    1179   1179       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   MOD_RES    1229   1229       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   VARIANT     512    512       A -> P.
FT                                /FTId=VAR_005299.
FT   VARIANT     723    723       MISSING (IN NIDDM).
FT                                /FTId=VAR_005301.
FT   VARIANT     971    971       G -> R.
FT                                /FTId=VAR_005300.
FT   VARIANT    1043   1043       S -> Y (IN NIDDM).
FT                                /FTId=VAR_005302.
FT   VARIANT    1095   1095       C -> Y (IN NIDDM).
FT                                /FTId=VAR_005303.
FT   CONFLICT    134    134       G -> GG (IN REF. 2).
FT   CONFLICT    362    362       S -> R (IN REF. 2).
FT   CONFLICT    384    384       P -> R (IN REF. 2).
SQ   SEQUENCE   1242 AA;  131590 MW;  3C0EFD9E32B3E64A CRC64;

  Query Match             4.7%;  Score 177.5;  DB 1;  Length 1242;
  Best Local Similarity   22.9%;  Pred. No. 0.048;
  Matches  117;  Conservative   57;  Mismatches  213;  Indels  123;  Gaps   22;

Qy      53 TDECGISLQCKDTLSLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEED 112
           | :|    :   ::|  |   | :|:   |:         :    |: | || |  :||
Db     403 TSDCLFPRRSSASVSGSPSDGGFISSDEYGSSPCDFRSSFRSVTPDSLGHTPPARGEEE- 461

Qy     113 DDEERAARRPGAGPP--KAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQV 170
                 | ||   | :|   |||     |    ||:|        |:  |
Db     462 --LSNYICMGGKGPSTLTAPNGHYILSRG---GNGHRCTPGTG------LGTSPALAGDE 510

Qy     171 PRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRG 230
           |   |:|    : |      |  :||  | ::||    :   ||
Db     511 AASAADLDNRFRKRTH------SAGTSPTITHQKTPSQSSVASIEEY------------ 551

Qy     231 TSTDSPCRRSTATQMAPPGGPPAAPPGGR--GHSHRDRIHYQADVRLEATEEIYLTPVQR 288
                   |:|  |  ||    |||  || |   :      |    | : : |::|
Db     552 -----------TEMM-PAYPPGGGSGGRLPGHRHSAFV----PTRSYPEEGLEMHPLER 594

Qy     289 PPDAAEPTSAFLPPTESRMSVSSDPDPAAYPSTAGR---------PHPSISEEEEGFDCL 339
               | |: |   :  | :|    |   |  :||           |:|  ::  : :
Db     595 RGGHHRPDSSTLHTDDGYMPMS----PGVAPVPSGRKGSGDYMPMSPKSVSAPQQIINPI 650

Qy     340 -SSPERAEP-------PGGGWRGSLGEPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQL 391
            |:| :|       | ||    :|  |    :| ::  |  ||  :    |   |: :
Db     651 RRHPQRVDPNGYMMMSPSGGCSPDIGGGPSSSSSSSNAVPSGTSYGKLWTNGVGGHHSHV 710

Qy     392 ----------ELVSLRPCFGDYSDES---DSATVYDNCASVSSPYESAIGEEYEEAPRPQ 438
                       | || ||| : |   || |       ||| :   |    | |: :
Db     711 LPHPKPPVESSGGKLLPCTGDYMNMSPVGDSNT--------SSPSDCYYG---PEDPQHK 759

Qy     439 P-------PACLSEDSTPDEPDVHFSKKFLNV-FMSGR----------SRSSSAESFGLF 480
           |       |        | ||:     : | :   |||          | |:|::| |
Db     760 PVLSYYSLPRSFKHTQRPGEPEEGARHQHLRLSTSSGRLLYAATADDSSSSTSSDSLGGG 819

Qy     481 SCIINGEEQEQT-----HRAIFRFVPRHED 505
           |   |   | :     |: :   :||  |
Db     820 YC---GARLEPSLPHPHHQVLQPHLPRKVD 846


RESULT 13
HLES_DROME
ID   HLES_DROME     STANDARD;      PRT;  1077 AA.
AC   Q02308;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hairless protein.
GN   H.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92387549; PubMed=1516831;
RA   Bang A.G., Posakony J.W.;
RT   "The Drosophila gene Hairless encodes a novel basic protein that
RT   controls alternative cell fates in adult sensory organ development.";
RL   Genes Dev. 6:1752-1769(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93041287; PubMed=1419850;
RA   Maier D., Stumm G., Kuhn K., Preiss A.;
RT   "Hairless, a Drosophila gene involved in neural development, encodes
RT   a novel, serine rich protein.";
RL   Mech. Dev. 38:143-156(1992).
CC   -!- FUNCTION: IS A POTENT ANTAGONIST OF NEUROGENIC GENE ACTIVITY
CC       DURING SENSORY ORGAN DEVELOPMENT. THE EXPRESSION OF DISTINCT CELL
CC       FATES BY THE TRICHOGEN (SHAFT) / TORMOGEN (SOCKET) SISTER CELL
CC       PAIR DEPENDS ON THE LEVEL OF H ACTIVITY. A CERTAIN THRESHOLD LEVEL
CC       OF H ACTIVITY IS REQUIRED, BELOW WHICH BOTH SISTER CELLS ADOPT THE
CC       TORMOGEN FATE.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: OVARY, EMBRYOS, LARVAL AND PUPAL IMAGINAL
CC       DISCS.
CC   -!- SIMILARITY: CONTAINS 1 PRD MOTIF.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M95192; AAA28607.1; ALT_INIT.
DR   EMBL; X67239; CAA47664.1; -.
DR   FlyBase; FBgn0001169; H.
KW   Developmental protein; Nuclear protein; DNA-binding.
FT   DOMAIN      115    123       THR-RICH.
FT   DOMAIN      642    648       POLY-SER.
FT   DOMAIN      879    891       POLY-ALA.
FT   DOMAIN      937    946       POLY-ALA.
FT   DOMAIN      964    974       ALA-RICH.
FT   DOMAIN      979   1008       HIS/PRO-RICH (PRD MOTIF).
FT   CONFLICT    151    151       S -> A (IN REF. 2).
FT   CONFLICT    702    703       QH -> LL (IN REF. 2).
FT   CONFLICT    891    891       A -> R (IN REF. 2).
FT   CONFLICT    964    967       AAVA -> RLLP (IN REF. 2).
FT   CONFLICT    974    974       MISSING (IN REF. 2).
SQ   SEQUENCE   1077 AA;  111039 MW;  A94BF1A27579E2F1 CRC64;

  Query Match             4.6%;  Score 174.5;  DB 1;  Length 1077;
  Best Local Similarity   21.5%;  Pred. No. 0.059;
  Matches  123;  Conservative   55;  Mismatches  230;  Indels  165;  Gaps   24;

Qy       6 SGGLGGGAA-------SPPAASPFLG-------LHIASPPNFRLTHDISLEEFEDEDLSE 51
Db     537 SDRIGGGLSHALTHKVSPPSSATAAGRLVEYHTQHVS--PRKRI-----LREFEKVSL-- 587

Qy      52 ITDECGISLQCKDTLSLRPPRAGLLSAGGGGAGSRLQ--------AEMLQMDLIDATGDT 103
Db     588 -----------EDNGCVNNGSGGASSGGAGGKRSRAKGTSTSSPAGKASPMNLAPPQG-- 634

Qy     104 PGAEDDEEDDDEERAARRPGAGPPKAESGQEPA---------------------SRGQGQ 142
Db     635 -----------------KPSPSPGSSSSSTSPATLSTQPTRLNSSYSIHSLLGGSSGSGS 677

Qy     143 SQGQSQGPGSGD---------------TYRPKRPTTLNLFPQVPRSQDTLNNNSLGKKHS 187
Db     678 SSSSSSGKKCGDHPAAIISNVHHPQHSMYQPSSSSYPRALLTSPKSPDVSGSNGGGGKSP 737

Qy     188 WQDRVSRSSSPLKTGEQTP---------------PHEHICLSEEL-PPQSGPA-PTTDRG 230
Db     738 SHTGTKKRSPPYSAGSPVDYGHSFYRDPYAGAGRPSTSGSASQDLSPPRSSPASPATTPR 797

Qy     231 T--STDSPCRRSTATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQR 288
Db     798 TVPKKTASIRREFASPSASSSSCPS--PGDRSASPPERRHMQQQPHLQRSSPLHYYMYPP 855

Qy     289 PP------DAAEPTSAFLPPTESRMSVSSDPDPAAYPSTAGRPHPSI------------- 329
Db     856 PPQVNGNGSAGSPTSA--PPTSN--SSAAAVAAAAAAAAAYIPSPSIYNPYISTLAALRH 911

Qy     330 -----SEEEEGFDCLSSPERAEPPGGGWRGSLGEPPPPPRASLSSDTSALSYDSVKYTLV 384
Db     912 NPLWMHHYQTGASPLLSPH--PQPGGS--AAAAAAAAAARLSPQSAYHAFAYNGV----- 962

Qy     385 VDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV-SSPYESAIGEEYEEAPRPQPPACL 443
Db     963 ---GAAVAAAAAAAAFG---QPAPSPHTHPHLAHPHQHPHPAALTTHHSPAHLATPKL-- 1014

Qy     444 SEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAES 476
Db    1015 -TDSSTDQMSATSSHRTASTSPSSSSASASSSA 1046


RESULT 14
SYN1_CANFA
ID   SYN1_CANFA     STANDARD;      PRT;   415 AA.
AC   O62732;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Synapsin I (Fragment).
GN   SYN1.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shibuya H., Liu P.-C., O'Brien D.P., Johnson G.S.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: NEURONAL PHOSPHOPROTEIN THAT COATS SYNAPTIC VESICLES,
CC       BINDS TO THE CYTOSKELETON, AND IS BELIEVED TO FUNCTION IN THE
CC       REGULATION OF NEUROTRANSMITTER RELEASE (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SYNAPSE.
CC   -!- SIMILARITY: BELONGS TO THE SYNAPSIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF049588; AAC05207.1; -.
DR   HSSP; P17599; 1AUX.
DR   InterPro; IPR001359; Synapsin.
DR   Pfam; PF02750; Synapsin_C; 1.
DR   PRINTS; PR01368; SYNAPSIN.
DR   PROSITE; PS00415; SYNAPSIN_1; PARTIAL.
DR   PROSITE; PS00416; SYNAPSIN_2; PARTIAL.
KW   Synapse; Phosphorylation; Neurone; Repeat; Actin-binding.
FT   NON_TER       1      1
FT   DOMAIN       <1    148       C (ACTIN-BINDING AND SYNAPTIC-VESICLE
FT                                 BINDING).
FT   DOMAIN      149    383       D (PRO-RICH LINKER).
FT   DOMAIN      384   >415       E.
FT   NON_TER     415    415
SQ   SEQUENCE   415 AA;  43388 MW;  E988E2026FC5361B CRC64;

  Query Match             4.5%;  Score 170.5;  DB 1;  Length 415;
  Best Local Similarity   23.6%;  Pred. No. 0.033;
  Matches   77;  Conservative   27;  Mismatches  115;  Indels  107;  Gaps   14;

Qy      82 GAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAARR--PGAGPPKAESGQEPASRG 139
Db     157 GRGSHSQT--------PSPGALPLGRQTSQQPSGPPAQQRPPPQGGPP--QPGPGPQRQG 206

Qy     140 ---------QGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNNNSLGKKHSWQD 190
Db     207 PPLQQRPPPQGQQHLSGLGPPAGSPL-PQRLPSPTSAPQQPVSQAQPLSQAQG------- 258

Qy     191 RVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRRSTATQMAPPGG 250
Db     259 ---RQSRPVAGGPGAPPAAR-------PPAS---PSPQRQAGPPQATRQTSVSGQAPPKA 305

Qy     251 PPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAFL---------P 301
Db     306 -SGAPPSGQQRQ--------------------GPPQKPPGPAGPTRQASQAGPMPRTGP 343

Qy     302 PTESRMSVSSDPDPAAYP----------------STAGRPHPSISEEEEGFDCLSSPER 344
Db     344 PT-TQQPRPSGPGPAGRPAKPQLAQKPSQDVPSPATAGGPPHPQLNKSQSLTNAFNLPE- 401

Qy     345 AEPPGGGWRGSLGEPPPPPRASLSSD 370
Db     402 --------------PAPPRPSLSQD 412


RESULT 15
SLP1_CLOTM
ID   SLP1_CLOTM     STANDARD;      PRT;  1664 AA.
AC   Q06852;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
```

```
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer
DE   protein 1).
GN   OLPB.
OS   Clostridium thermocellum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1515;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCIB 10682;
RX   MEDLINE=93209931; PubMed=8458832;
RA   Fujino T., Beguin P., Aubert J.-P.;
RT   "Organization of a Clostridium thermocellum gene cluster encoding the
RT   cellulosomal scaffolding protein CipA and a protein possibly involved
RT   in attachment of the cellulosome to the cell surface.";
RL   J. Bacteriol. 175:1891-1899(1993).
CC   -!- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.
CC   -!- SUBCELLULAR LOCATION: Cell wall.
CC   -!- SIMILARITY: CONTAINS 4 S-LAYER HOMOLOGY (SLH) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67506; CAA47841.1; -.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF00395; SLH; 3.
DR   PROSITE; PS01072; SLH_DOMAIN; 2.
KW   Cell wall; S-layer; Signal; Repeat.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29   1664       CELL SURFACE GLYCOPROTEIN 1.
FT   DOMAIN       36    763       4 X 156 AA APPROXIMATE REPEATS.
FT   REPEAT       36    191       1.
FT   REPEAT      207    363       2.
FT   REPEAT      409    565       3.
FT   REPEAT      607    763       4.
FT   DOMAIN      771   1377       APPROXIMATE TANDEM REPEATS OF
FT                                T-P-S-D-E-P.
FT   DOMAIN     1378   1449       GLY/PRO/SER/THR-RICH.
FT   DOMAIN     1453   1494       SLH 1 (INCOMPLETE).
FT   DOMAIN     1495   1565       SLH 2.
FT   DOMAIN     1566   1625       SLH 3.
FT   DOMAIN     1626   1646       SLH 4 (INCOMPLETE).
SQ   SEQUENCE   1664 AA;  178194 MW;  5F396695BA9FE74B CRC64;

  Query Match             4.5%;  Score 170.5;  DB 1;  Length 1664;
  Best Local Similarity   22.8%;  Pred. No. 0.15;
  Matches   86;  Conservative   39;  Mismatches  202;  Indels   51;  Gaps   11;

Qy      98 DATGDTPGAEDDEEDDDEERAARRPGAG--PPKAESGQEPASRGQGQSQ-------GQSQ 148
           |   | |   |:    ||   :  |     |  :|: :||        :         |
Db     862 DTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSD 921

Qy     149 GPGSGDTYRPKRPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPH 208
           |   |    |    | :  |: |   || :           |  : |  |  : | ||
Db     922 EPTPSDEPTPSDEPTPSETPEEPIPTDTPS----------DEPTPSDEPTPSDEPTP-- 968

Qy     209 EHICLSEELPPQSGPAPTTDRGTSTDSPCRRSTATQMAPPGGPPAAPPGGRGHSHRDRIH 268
             |:|  |   | | :|  | :| |    |  :  |   |   |      :  |
Db     969 -----SDEPTPSDEPTP-SDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSD--- 1019

Qy     269 YQADVRLEATEEIYLTPVQRPPDAAEPTSAFLP-----------PTESRMSVSSD-PDP 315
           :    | |     ||   |  : |||:  |            || |     || | |
Db    1020 -EPTPSDEPTPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTP 1078

Qy     316 AAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEPPPPPRASLSSDTSALS 375
           :  |: :  | || : ||     : :   :: |      :  : | |      ||    |
Db    1079 SDEPTPSDEPTPSETPEEP----IPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1134

Qy     376 YDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASVSSPYESAIGEEYEEAP 435
            :: :  :  | :      |  |  |    || | |     :| |    :  : |
Db    1135 -ETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSETPEEPIPTDTPSDEP 1193

Qy     436 RPQPPACLSEDSTP-DEP 452
           |      |:: || |||
Db    1194 TPSDEPTPSDEPTPSDEP 1211


Search completed: December 24, 2002, 08:49:40
Job time : 46 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: December 24, 2002, 08:48:50 ; Search time 101 Seconds
(without alignments)
1450.490 Million cell updates/sec

Title: US-09-966-561-2
Perfect score: 3792
Sequence: 1 MAERESGGLGGGAASPPAAS.....QQFYKQFVEYTCPTEDIYLE 711

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3541	93.4	707	11	Q925J8	Q925j8 mus musculu
2	1100	29.0	455	4	Q96G62	Q96g62 homo sapien
3	246.5	6.5	315	5	P91045	P91045 caenorhabdi
4	211	5.6	59	11	Q9CUY3	Q9cuy3 mus musculu
5	210	5.5	647	10	Q8S148	Q8s148 oryza sativ
6	210	5.5	950	11	Q923A9	Q923a9 mus musculu
7	210	5.5	2157	11	Q9Z1R1	Q9z1r1 mus musculu
8	200.5	5.3	1217	4	Q9ULL5	Q9ull5 homo sapien
9	190.5	5.0	1735	4	Q9HBL0	Q9hbl0 homo sapien
10	190	5.0	2157	4	O95875	O95875 homo sapien
11	190	5.0	2157	4	Q96QC6	Q96qc6 homo sapien
12	184	4.9	1715	6	Q9GLM4	Q9glm4 bos taurus
13	183.5	4.8	640	4	O60593	O60593 homo sapien
14	182	4.8	1634	5	Q9XW25	Q9xw25 caenorhabdi
15	180.5	4.8	1251	6	Q28224	Q28224 cercopithec
16	179.5	4.7	4957	4	O14687	O14687 homo sapien
17	179.5	4.7	5262	4	O14686	O14686 homo sapien
18	179	4.7	623	5	Q9W4A2	Q9w4a2 drosophila
19	178	4.7	926	5	Q9W3G1	Q9w3g1 drosophila
20	177.5	4.7	1240	13	P79773	P79773 gallus gall
21	177	4.7	1240	12	Q9DWH8	Q9dwh8 rat cytomeg
22	176.5	4.7	917	4	Q9P1Y5	Q9ply5 homo sapien
23	173.5	4.6	645	4	Q96EX0	Q96ex0 homo sapien
24	173.5	4.6	1110	5	Q09493	Q09493 caenorhabdi
25	173	4.6	304	4	Q9UBP9	Q9ubp9 homo sapien
26	172	4.5	857	3	Q9P7E8	Q9p7e8 schizosacch
27	171.5	4.5	619	5	Q960W5	Q960w5 drosophila
28	171.5	4.5	1077	5	Q9VDK1	Q9vdk1 drosophila
29	171.5	4.5	1279	5	O97005	O97005 leishmania
30	171.5	4.5	1307	10	Q9LVN1	Q9lvn1 arabidopsis
31	171	4.5	302	11	Q9CRV4	Q9crv4 mus musculu
32	170.5	4.5	440	4	Q8TF74	Q8tf74 homo sapien
33	170.5	4.5	440	4	Q8TE44	Q8te44 homo sapien
34	170.5	4.5	487	11	Q8VDA4	Q8vda4 rattus norv
35	170.5	4.5	812	12	Q9DWG7	Q9dwg7 rat cytomeg
36	170.5	4.5	1056	4	O95621	O95621 homo sapien
37	170.5	4.5	1182	4	Q99495	Q99495 homo sapien
38	170.5	4.5	1190	4	Q99621	Q99621 homo sapien
39	170.5	4.5	1201	11	Q8VHK1	Q8vhk1 mus musculu
40	170.5	4.5	2158	11	Q9WU13	Q9wu13 rattus norv
41	170	4.5	309	11	Q9CYD2	Q9cyd2 mus musculu
42	169.5	4.5	1044	4	Q9P206	Q9p206 homo sapien
43	169.5	4.5	2112	5	Q9VEL9	Q9vel9 drosophila
44	169.5	4.5	3105	12	Q8UZI9	Q8uzi9 cercopithic
45	169	4.5	1386	4	Q9C0A3	Q9c0a3 homo sapien

ALIGNMENTS

RESULT 1
Q925J8
ID Q925J8 PRELIMINARY; PRT; 707 AA.
AC Q925J8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein kinase mitogen-activated 8 interacting protein.
GN MAPK8IP OR PRKM8IP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ILS;
RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;
RT "High-Throughput Sequence Identification of Gene Coding Variants
RT within Alcohol-Related QTLs.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF332076; AAK56104.1; -.
DR MGD; MGI:1309464; Mapk8ip.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00640; PID; 1.
DR Pfam; PF00018; SH3; 1.
DR PROSITE; PS01179; PID; 1.
DR PROSITE; PS50002; SH3; 1.
KW Kinase.
SQ SEQUENCE 707 AA; 77381 MW; FDA53D891B6F5AA6 CRC64;

Query Match 93.4%; Score 3541; DB 11; Length 707;
Best Local Similarity 93.5%; Pred. No. 1.3e-250;
Matches 666; Conservative 13; Mismatches 27; Indels 6; Gaps 3;

```
Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
           |||||| ||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAERES-GLGRGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 59
```

```
Qy      61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
           ||||||||||||||||||  | ||||||||||||||| |||||||||||::|:| ||
Db      60 QCKDTLSLRPPRAGLLSAGSSGSAGSRLQAEMLQMDLIDAAGDTPGAEDDEEEEDDELAA 119

Qy     120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
           :||| ||||||| |:|| |    |||| | ||||||||||||||||||||||||||||||
Db     120 QRPGVGPPKAESNQDPAPR----SQGQGPGTGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 175

Qy     180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
           ||||||||||||||||||||||||||||||||:|||||  | || ||||||||||||
Db     176 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSDELPPQGSPVPTQDRGTSTDSPCRR 235

Qy     240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
           | ||||||| |||| ||||||||||||||||||||||||||||||||| ||||| |
Db     236 SAATQMAPPSGPPATAPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDPAEPTSTF 295

Qy     300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
           :|||||||||||||||||  ||||||||||||:|||||||||||||||||||||||||||
Db     296 MPPTESRMSVSSDPDPAAYSVTAGRPHPSISEEDEGFDCLSSPERAEPPGGGWRGSLGEP 355

Qy     360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     356 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASA 415

Qy     420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
           ||||||||||||||:|:|| ||||||||||||||||||||||||||||||||||||||||
Db     416 SSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 475

Qy     480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
           |||:||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     476 FSCVINGEEHEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 535

Qy     540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
           ||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db     536 PAYYAIEVTKEPEHMAALAKNSDWIDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 595

Qy     600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
           |||||||||||||||||||||||||: ||||||||||||||||||||||||||||||||
Db     596 TVHFNPPSSCVLEISVRGVKIGVKADDALEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 655

Qy     660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     656 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 707
```

RESULT 2
Q96G62
ID Q96G62 PRELIMINARY; PRT; 455 AA.
AC Q96G62;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to mitogen-activated protein kinase 8 interacting protein
DE 2 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC009940; AAH09940.1; -.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00640; PID; 1.
DR Pfam; PF00018; SH3; 1.
DR PROSITE; PS01179; PID; 1.
DR PROSITE; PS50002; SH3; 1.
KW Kinase.
FT NON_TER 1 1

```
SQ   SEQUENCE   455 AA;  50405 MW;  F10CD3C4C6A8E522 CRC64;

  Query Match             29.0%;  Score 1100;  DB 4;  Length 455;
  Best Local Similarity   35.7%;  Pred. No. 2.4e-72;
  Matches  252;  Conservative   64;  Mismatches   95;  Indels  294;  Gaps   14;

Qy      26 HIASPPNFRLTHDISLEEFEDEDLSEITDECGISL-----QC-KDTLSLRPPRAGLLSAG 79
           |  |||  |  |||||||:||||||||:||: |      | ||:|||
Db      26 HSLSPPGCRPPQDISLEEFDDEDLSEITDDCGLGLSYDSDHCEKDSLSL---------- 74

Qy      80 GGGAGSR------LQAEMLQMDLIDATGDTPGAEDDEEDDDEERAARRPGAGPPKAESGQ 133
             |   :       | :  : ::||        |:|||||:||  |
Db      75 --GRSEQPHPICSFQDDFQEFEMID-----DNEEEDEEDDEEEEDAE------------ 114

Qy     134 EPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVS 193

Db     115 ------------------------------------------------------------ 114

Qy     194 RSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRRSTATQMAPPGGPPA 253
                                                  ||                  |
Db     115 ----------------------------------------DS-----------------A 117

Qy     254 APPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAFLPPTESRMSVSSDP 313
             |||||
Db     118 GSPGGRGTG---------------------------------------------------- 126

Qy     314 DPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEPPPPPRASLSSDTSA 373
                                                      |  ||       ::
Db     127 ------------------------------------------PSAPR------DAS 134

Qy     374 LSYDSVKYTLVVDEHAQLELVSLRPCFG-DYSDESDSATVYDNCASVSSPYESAIGEEYE 432
           | ||:|||||||||| |||||||| | |  :  | ||                  ||  |
Db     135 LVYDAVKYTLVVDEHTQLELVSLRRCAGLGHDSEEDSG----------------GEASE 177

Qy     433 EAPRPQPPACL------SEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLFSCIING 486
           |    :  | |       | |::|| ||: |||||||||::  ||||| ||||||||::||
Db     178 E----EAGAALLGGGQVSGDTSPDSPDLTFSKKFLNVFVNSTSRSSSTESFGLFSCLVNG 233

Qy     487 EEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFPAYYAIE 546
           ||:|||||:|||:||| |||||:||||:||| : :|:|:  :||||| ||||||:||
Db     234 EEREQTHRAVFRFIPRHPDELELDVDDPVLVEAEEDDFWFRGFNMRTGERGVFPAFYAHA 293

Qy     547 VTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRLTVHFNPP 606
           |    : :   ::  ||::| |:|||||:|| |:|| :|||||||||| |:|||  ||
Db     294 VPGPAKDLLGSKRSPCWVERFDVQFLGSVEVPCHQGNGILCAAMQKIATARKLTVHLRPP 353

Qy     607 SSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGFITKHPAD 666
           :|| ||||:|||||: :     :     :||||||:||||||| ||:|: |||||||||
Db     354 ASCDLEISLRGVKLSLSGGGPE---FQRCSHFFQMKNISFCGCHPRNSCYFGFITKHPLL 410

Qy     667 HRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
           |||||||||:| : :|:|||||| ::|:: : | |||||||||
Db     411 SRFACHVFVSQESMRPVAQSVGRAFLEYYQEHLAYACPTEDIYLE 455
```

RESULT 3
P91045
ID P91045 PRELIMINARY; PRT; 315 AA.
AC P91045;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE C13A10.3 protein.
GN C13A10.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;

```
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Wilson R., Pauley A., Maggi L.;
RT   "The sequence of C. elegans cosmid C13A10.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR   EMBL; U80841; AAB37940.1; -.
DR   InterPro; IPR000050; PID_domain.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00018; SH3; 1.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00462; PTB; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   SH3 domain.
SQ   SEQUENCE   315 AA;  36385 MW;  B8572746211CFAAC CRC64;

  Query Match             6.5%;  Score 246.5;  DB 5;  Length 315;
  Best Local Similarity   36.8%;  Pred. No. 3.9e-10;
  Matches   56;  Conservative   27;  Mismatches   62;  Indels    7;  Gaps    3;

Qy    475 ESFGLFSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTG 534
          :| |: ||    : |  |||    | ||| ||| ||: | : |:  |:|:|    |:|||
Db     31 DSSGVSSC-TTSDSQNPTHRVQSAFHPRHPDELLLEIGDAVHVDRTADDHWSYGTNLRTG 89

Qy    535 ARGVFPAYYAIEVTKEPE-HMAALAKNSDWV-----DQFRVKFLGSVQVPYHKGNDVLCA 588
           |:|||    |:    |  : ||  |:  :     | | :  | |::| :|||||||
Db     90 QSGIFPASIVCEIDLVEEICLGALPTNATKILSGDRDTFYLTMLASIEVAHHKGNDVLTQ 149

Qy    589 AMQKIATTRRLTVHFNPPSSCVLEISVRGVKI 620
          || |: :  : :     | : ::||| ||: :
Db    150 AMNKVLSMYKNSEEIIVPQTVLMEISFRGIHV 181


RESULT 4
Q9CUY3
ID   Q9CUY3      PRELIMINARY;      PRT;    59 AA.
AC   Q9CUY3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   3230402N03Rik protein (Fragment).
GN   MAPK8IP2 OR 3230402N03RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK013573; BAB28912.1; -.
DR   MGD; MGI:1926555; Mapk8ip2.
DR   InterPro; IPR000050; PID_domain.
DR   Pfam; PF00640; PID; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   59 AA;  6877 MW;  35C6B6E84D02FC2C CRC64;

  Query Match             5.6%;  Score 211;  DB 11;  Length 59;
  Best Local Similarity   62.7%;  Pred. No. 1.7e-08;
  Matches   37;  Conservative   11;  Mismatches   11;  Indels    0;  Gaps    0;

Qy    653 NNKYFGFITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          |: |||||||||   ||||||||||::| : :| |||||| ::|:: : : |||||||||
Db      1 NSCYFGFITKHPLLSRFACHVFVSQESMRPVARSVGRAFLEYYQEHLAFACPTEDIYLE 59


RESULT 5
Q8S148
ID   Q8S148      PRELIMINARY;      PRT;    647 AA.
AC   Q8S148;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   VsaA-like protein.
GN   P0042A10.17.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
RT   clone:P0042A10.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003343; BAB90075.1; -.
SQ   SEQUENCE   647 AA;  62062 MW;  C27F3FA0A77D1EC3 CRC64;

  Query Match             5.5%;  Score 210;  DB 10;  Length 647;
  Best Local Similarity   23.2%;  Pred. No. 4.7e-07;
  Matches  121;  Conservative   25;  Mismatches  183;  Indels  192;  Gaps   22;

Qy     10 GGGAASPPAASPFLGLHIASP---------PNFRLTHDISLEEFEDEDLSEITDECGISL 60
          |||  :||  || :|    :|          |:                | |  |   | |
Db     57 GGGGYNPP--SPSIGTSPTTPGGGGGYTPTPS-------DTPPSPSSDTSPSTPGGGCSS 107

Qy     61 Q----------CKDTLSLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGA---- 106
                    ||    |        ||||  |   ::       | :  |||
Db    108 SPTPCDAPPSPSSDTSPTTP--------GGGGGYSPTPSDTPPSPSSDTSPTTPGGGGGY 159

Qy    107 ----EDDEEDDDEERAARRPGAG----------PPKAESGQEPASRGQGQSQGQSQGPGS 152
              |       : :   || |          ||   |   | :            ||
Db    160 TPTPSDAPPSPSSDTSPTTPGGGGGYTPTPSDAPPSPSSDTSPTT-----------PGG 207

Qy    153 GDTYRPKRPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTG---------E 203
          | | |    |  | |                        |  |||   |
Db    208 GGGYTPTPSDT----PPSP--------------------SSGSSPTTPGGGGGYTPTPS 242

Qy    204 QTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPC----RRSTATQMAPPGG--------- 250
           |||          | || : ||  | || :||      |: |    |||
Db    243 DTPP----------SPSSGSSRTTPGGCSTPTPCGTPPAPSSGTSPTTPGGSYYPPTPSI 292

Qy    251 ---PPA-------APPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTS--- 297
             ||:         |||   |                        ||   ||  :  ||
```

```
Db    293 GDVPPSPSSDTSPTTPGGGSPS--------------------TPCDTPPSPSSGTSPTT 331

Qy    298 ---AFLPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRG 354
Db    332 PGGGYYPPTPSVGDVPPSPASGTSPSTPGSGGYSPSTPCSAPPSPSSGTSPTTPGGGYSP 391

Qy    355 S--LGEPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATV 412
Db    392 STPCNAPPSP-----SSDTSPTT----------------------PGGGNYPPAPTIGNV 424

Qy    413 YDNCASVSSPYESAIGEEYEEAPRP-QPPACLSEDSTPDEP 452
Db    425 PPSPSSGTSP--STPGGGCSSSPTPCDAPPSPSSDTSPTTP 463


RESULT 6
Q923A9
ID   Q923A9      PRELIMINARY;      PRT;   950 AA.
AC   Q923A9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Similar to RIKEN cDNA 3110039B05 gene (Fragment).
GN   BAT2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC006664; AAH06664.1; -.
DR   MGD; MGI:1915467; Bat2.
FT   NON_TER       1       1
SQ   SEQUENCE   950 AA;  99726 MW;  E41DE9AB4A66D976 CRC64;

  Query Match             5.5%;  Score 210;  DB 11;  Length 950;
  Best Local Similarity   22.9%;  Pred. No. 7.7e-07;
  Matches  123;  Conservative   33;  Mismatches  200;  Indels  182;  Gaps   21;

Qy      3 ERESGGLGGG--------AASPPAASPFLGLHIASPPNFRL---THDISLEEFED----- 46
Db     50 ERENAARGADGKPPSLTLAASTPGPEETLTAATVPPPPRRTAAKSPDLSNQNSDQANEEW 109

Qy     47 EDLSEITDECGISLQCKDTLSLRPPRA-------GLLSAGGGGAGSRLQAEMLQMDLIDA 99
Db    110 ETASESSDFASERRGDKET----PPAALMTSKAVGTPGANAGGAGPGISA-MSRGDLSQR 164

Qy    100 TGDTPGAEDDEEDDDEERAARRPGAGPPKAESGQEPASRGQGQSQGQSGPGSGDTYRPK 159
Db    165 AKDLSKRSFSSQRPGMDRQNRRPGTGGKTGSGG--GSSGGGGAGPGGRTGPGRGD----- 217

Qy    160 RPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPP 219
Db    218 ------------------------KRSWPSPKNRSRPP----EERPP------GLPLPP 242

Qy    220 QSGPAPTTDRGTSTDSPCRRSTA-------------TQMAPPGGPPAAPPGGRGHSHRDR 266
Db    243 ---PPPSSSAVFRLDQVIHSNPAGIQQALAQLSSRQGNVTAPGGHPRPKPG--------- 290

Qy    267 IHYQADVRLEATEEIYLTPVQRPPDAA-EPTSAFLPPTESRMSVSSDP---DPA-AYPST 321
Db    291 -----------------PPQAPQGSSPRPPTRYDPPRAS-SAISSDPHFEEPGPMVRGV 331

Qy    322 AGRPHPSISEEEEGFDCLSSPERAEPP--------------GGGWRGSLGEPPPPPRAS 366
Db    332 GGTPRDS-----AGVNPFPPKRRERPPRKPELLQEETVPASHSSGFLGSKPEVPGPQEES 386

Qy    367 LSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCF----------GDYSDESDSATVYDNC 416
Db    387 RDSGTEALTPHIWNRLHTATSRKSYQPGSIEPWMEPLSPFEDVAGTEMSQSDSGVDLSGD 446

Qy    417 ASVSS--------------------------------PYESAIGEEYEEAPRPQPPA 441
Db    447 SQVSSGPCSQRSSPDGGLKGSAEGPPKRPGGPSPLNAVPGESASGSEPSEPPRRRPPA 504


RESULT 7
Q9Z1R1
ID   Q9Z1R1      PRELIMINARY;      PRT;  2157 AA.
AC   Q9Z1R1;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BAT2.
GN   NFKBIL1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rowen L., Qin S., Madan A., Abbasi N., James R., Dickhoff R.,
RA   Shaffer T., Ratcliffe A., Loretz C., Lasky S., Hood L.;
RT   "Sequence of the mouse major histocompatibility class III region.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF109719; AAC82480.1; -.
DR   MGD; MGI:1340031; Nfkbil1.
SQ   SEQUENCE   2157 AA;  229072 MW;  35B21F6B7C971F40 CRC64;

  Query Match             5.5%;  Score 210;  DB 11;  Length 2157;
  Best Local Similarity   22.9%;  Pred. No. 2.3e-06;
  Matches  123;  Conservative   33;  Mismatches  200;  Indels  182;  Gaps   21;

Qy      3 ERESGGLGGG--------AASPPAASPFLGLHIASPPNFRL---THDISLEEFED----- 46
Db   1258 ERENAARGADGKPPSLTLAASTPGPEETLTAATVPPPPRRTAAKSPDLSNQNSDQANEEW 1317

Qy     47 EDLSEITDECGISLQCKDTLSLRPPRA-------GLLSAGGGGAGSRLQAEMLQMDLIDA 99
Db   1318 ETASESSDFASERRGDKET----PPAALMTSKAVGTPGANAGGAGPGISA-MSRGDLSQR 1372

Qy    100 TGDTPGAEDDEEDDDEERAARRPGAGPPKAESGQEPASRGQGQSQGQSGPGSGDTYRPK 159
Db   1373 AKDLSKRSFSSQRPGMDRQNRRPGTGGKTGSGG--GSSGGGGAGPGGRTGPGRGD----- 1425

Qy    160 RPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPP 219
Db   1426 ------------------------KRSWPSPKNRSRPP----EERPP------GLPLPP 1450

Qy    220 QSGPAPTTDRGTSTDSPCRRSTA-------------TQMAPPGGPPAAPPGGRGHSHRDR 266
Db   1451 ---PPPSSSAVFRLDQVIHSNPAGIQQALAQLSSRQGNVTAPGGHPRPKPG--------- 1498

Qy    267 IHYQADVRLEATEEIYLTPVQRPPDAA-EPTSAFLPPTESRMSVSSDP---DPA-AYPST 321
Db   1499 -----------------PPQAPQGSSPRPPTRYDPPRAS-SAISSDPHFEEPGPMVRGV 1539

Qy    322 AGRPHPSISEEEEGFDCLSSPERAEPP--------------GGGWRGSLGEPPPPPRAS 366
Db   1540 GGTPRDS-----AGVNPFPPKRRERPPRKPELLQEETVPASHSSGFLGSKPEVPGPQEES 1594

Qy    367 LSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCF----------GDYSDESDSATVYDNC 416
Db   1595 RDSGTEALTPHIWNRLHTATSRKSYQPGSIEPWMEPLSPFEDVAGTEMSQSDSGVDLSGD 1654

Qy    417 ASVSS--------------------------------PYESAIGEEYEEAPRPQPPA 441
Db   1655 SQVSSGPCSQRSSPDGGLKGSAEGPPKRPGGPSPLNAVPGESASGSEPSEPPRRRPPA 1712


RESULT 8
Q9ULL5
ID   Q9ULL5      PRELIMINARY;      PRT;  1217 AA.
AC   Q9ULL5;
```

```
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA1205 protein (Fragment).
GN   KIAA1205.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=20039619; PubMed=10574462;
RA   Nagase T., Ishikawa K., Kikuno R., Hirosawa M., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XV.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:337-345(1999).
DR   EMBL; AB033031; BAA86519.1; -.
DR   InterPro; IPR000637; AT_hook.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR00929; ATHOOK.
DR   PRINTS; PR01217; PRICHEXTENSN.
FT   NON_TER       1       1
SQ   SEQUENCE   1217 AA;  130189 MW;  EC83FC3173861B73 CRC64;

  Query Match             5.3%;  Score 200.5;  DB 4;  Length 1217;
  Best Local Similarity   21.3%;  Pred. No. 5.3e-06;
  Matches  132;  Conservative   57;  Mismatches  217;  Indels  213;  Gaps   26;

Qy      11 GGAASP----------PAASPFLGLHIASPPNFRLTHDISLEEFE---DEDLSEITDECG 57
           ||  ||          | |  |    || |:    :  :|:|   ||| :::  :
Db     238 GGLTSPIFCSTKPKKLLKTSSFHLLRRRDPP-FQTPKKLYAQEYEFEADEDKADVPAD-- 294

Qy      58 ISLQCKDTLSLRPPR-AGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEE 116
                : | | |   |:|:       | :  :  :  ||     || :       :
Db     295 --------IRLNPRRLPDLVSS------CRSRPALSPLGDIDFCPPNPGPDGPRRRGRKP 340

Qy     117 RAARRPG-----------------AGPPKAESGQEPAS--RGQGQSQGQSQGPGSGDTY 156
             |:| |                 |||  | :  : |  ||:|: :|:      |
Db     341 TKAKRDGPPRPRGRPRIRPLEVPTTAGPASASTPTDGAKKPRGRGRGRGRKAEEAGGTRL 400

Qy     157 RPKRPTTLNLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSS----------------P- 198
           | :|  : |   ||:: : |  :| |   |  | |  ||                   |
Db     401 EPLKPLKIKL--SVPKAGEGLGTSS-GDAISGTDHNSLDSSLTREKIEAKIKEVEEKQPE 457

Qy     199 ------------LKTGEQTPPHEHICLSEEL-PPQSGPAPTTDRGTSTDSPCRRSTATQM 245
                      ||:|:: ||     |:  | ||:| |     ||     |   : |
Db     458 MKSGFMASFLDFLKSGKRHPPLYQAGLTPPLSPPKSVPPSVPARGLQPQPPA--TPAVPH 515

Qy     246 APPGGP-------PAAPPGGRG-------------------------HSHRDRIHYQADV 273
           || |          ||   | |                        | :    |:
Db     516 PPPSGAFGLGGALEAAESEGLGLGCPSPCKRLDEELKRNLETLPSFSSDEEDSVAKNRDL 575

Qy     274 RLEATEEIYL---TPVQRPPDAAEPTSAFLPPTESRMSVSSDPDPAAYPS--TAGRPHPS 328
           :   :  |      |:  | | : |     ||     :|   |     ||  : | | |
Db     576 QESISSAISALDDPPLAGPKDTSTPDG---PPLAPAAAVPGPPPLPGLPSANSNGTPEPP 632

Qy     329 ISEEEEGFDCLSSPERAEPPGGGWRGSLGEPPPPPRASLSSDTSALSYDSVKYTLVVDEH 388
           : ||:        |    ||    :     |||||: :| |
Db     633 LLEEK--------PPPTPPPAPTPQPQPPPPPPPPQPALPSPP----------------- 667

Qy     389 AQLELVSLRPCFGDYSDESDSATVYDNCASVSSPYESAIGEEYEEA----PRPQPPACLS 444
              ||:  |                 |||    :      |     | | |||
Db     668 ---PLVAPTP--------------------SSPPPPPLPPPPPPAMPSPPPPPPPAAAP 703

Qy     445 EDSTPDEP-------------DVHFSKKFLNVFMSGRSRSSSAESFGLFSCIINGEEQE 490
             : |:||              :| :||     : | :   : || |
Db     704 LAAPPEEPAAPSPEDPELPDTRPLHLAKKQETAAVCGETDEEAGESGG------------ 751

Qy     491 QTHRAIFR----FVPRHED 505
                |||    || | ||
Db     752 ---EGIFRERDEFVIRAED 767
```

```
RESULT 9
Q9HBL0
ID   Q9HBL0      PRELIMINARY;      PRT;  1735 AA.
AC   Q9HBL0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Tensin.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RX   MEDLINE=20480062; PubMed=11023826;
RA   Chen H., Ishii A., Wong W.K., Chen L.B., Lo S.H.;
RT   "Molecular characterization of human tensin.";
RL   Biochem. J. 351:403-411(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RA   Lo S.H., Chen L.B.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF225896; AAG33700.1; -.
DR   HSSP; P00519; 1AB2.
DR   InterPro; IPR000340; DS_phosphatase.
DR   InterPro; IPR000209; Peptidase_S8.
DR   InterPro; IPR000050; PID_domain.
DR   InterPro; IPR000980; SH2.
DR   Pfam; PF00017; SH2; 1.
DR   ProDom; PD000093; SH2; 1.
DR   SMART; SM00462; PTB; 1.
DR   SMART; SM00012; PTPc_DSPc; 1.
DR   SMART; SM00252; SH2; 1.
DR   PROSITE; PS50001; SH2; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
SQ   SEQUENCE   1735 AA;  185677 MW;  AC825375DD6CF87C CRC64;

  Query Match             5.0%;  Score 190.5;  DB 4;  Length 1735;
  Best Local Similarity   21.2%;  Pred. No. 4.5e-05;
  Matches  185;  Conservative   87;  Mismatches  318;  Indels  283;  Gaps   39;

Qy      12 GAASP-PAASPFLGLHIASPPNFR------LTHDISLEEFEDEDLSEITDECGISLQCKD 64
           |  ||    || | | ||  |  |       ::  : || |    |
Db     917 GVRSPVQCVSPELALTIALNPGGRPKEPHLHSYKEAFEEMEGTSPSS------------ 963

Qy      65 TLSLRPPRAGLLSAGGGGAGSRLQAEMLQMD-----LIDATGDTPGAEDDEEDDDE---- 115
             || :|: |   | | : | |  |:       |:  || |      |||: :
Db     964 -----PPPSGVRSP-PGLAKTPLSALGLKPHNPADILLHPTGVTRRRIQPEEDEGKVVVR 1017

Qy     116 ---------ERAARRPGAGP------PKAESGQEPASRGQGQSQGQSQGPGSGDTYRPK- 159
                   |  ||   |||      ||: |   ||::  |       |   |    |
Db    1018 LSEEPRSYVESVARTAVAGPRAQDSEPKSFSA--PATQAYGHEIPLRNGTLGGSFVSPSP 1075

Qy     160 ----------RPTTLNLFP------QVPRS------QDTLNNNSLGKKHSWQDRVSRSSS 197
                    |::  ||       | ||:       :    : |||:      |  :|||
Db    1076 LSTSSPILSADSTSVGSFPSGESSDQGPRTPTQPLLESGFRSGSLGQPSPSAQRNYQSSS 1135

Qy     198 PLKT-----------------------------GEQT----PPHEHICLSEELPP---- 219
           || |                             |  |     |   |  :    ||
Db    1136 PLPTVGSSYSSPDYSLQHFSSSPESQARAQFSVAGVHTVPGSPQARHRTVGTNTPPSPGF 1195

Qy     220 --------------------QSGPAPTTDRGTSTDSPCRRSTATQMAP----------- 247
                              ||  |   |::  ||   :  |  :|
Db    1196 GRRAINPSMAAPSSPSLSHHQMMGPPGTGFHGSTVSSPQSSAATTPGSPSLCRHPAGVYQ 1255

Qy     248 -----------PGGP-----------------------PAAPPGGR----------GHS 262
                      || |                       | :|  ||          |
Db    1256 VSGLHNKVATTPGSPSLGRHPGAHQGNLASGLHSNAIASPGSPSLGRHLGGSGSVVPGSP 1315
```

```
Qy     263 HRDR-IHYQADVRLEATEEIYLTPVQRPPDAAEPTSA--FLPPTESRMSVSSDPDPAAYP 319
           || : |              | ||  | |  :  :||  :  |:     ||    || ||
Db    1316 CLDRHVAYGG----------YSTPEDRRPTLSRQSSASGYQAPSTPSFPVS----PAYYP 1361

Qy     320 ---STAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSL-------GEPPPPPRASLSS 369
             | |  | |  :   :|    : ||:     |   |||       |:   |  : :||
Db    1362 GLSSPATSPSPDSAAFRQGSPTPALPEKRRMSVGDRAGSLPNYATINGKVASPVPSGMSS 1421

Qy     370 DTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASVSSPYESAIGE 429
           :   |  |  :||               | |  ||   :|    :  | |    :::
Db    1422 PSGG-STVSFSHTL-------------PDFSKYSMPDNSP---ETRAKVKFVQDTSKYW 1463

Qy     430 EYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLFSCIINGEEQ 489
             |  | |  | :|  |     :  |  |   :      ||   :      :    ::
Db    1464 YKPEISREQAIALL-KDQEPGAFIIRDSHSFRGAYGLAMKVSSPPPTI-----MQQNKKG 1517

Qy     490 EQTHRAIFRFV----PRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFPAYYAI 545
           : ||  :  |:    ||   :|:   ::|    | |  | :     :    : | |
Db    1518 DMTHELVRHFLIETGPRGV-KLKGCPNEPNFGSLSALVYQHSIIPLALPCKLVIPNRDPT 1576

Qy     546 EVTKEPEHMAALAKNSDWVDQ---FRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRLTVH 602
           : :|:    |     :| : |     | |: || :    |       |: | ||:  |
Db    1577 DESKDSSGPA--NSTADLLKQGAACNVLFINSVDMESLTGPQ----AISK-ATSETLAAD 1629

Qy     603 FNPPSSCV-LEISVRGVKIGVKADDSQEAKGNKCSHFFQ----LKNISFCGYHPKNNKY- 656
             | :: |  ::| :|: :            |:   ||:    |  ::||   |:  |:
Db    1630 PTPAATIVHFKVSAQGITL----------TDNQRKLFFRRHYPLNTVTFCDLDPQERKWM 1679

Qy     657 ----------FGFIT-KHPADHRFACHVFVSED 678
                    |||:  |  :    |||:|   |
Db    1680 KTEGGAPAKLFGFVARKQGSTTDNACHLFAELD 1712


RESULT 10
O95875
ID   O95875      PRELIMINARY;      PRT;  2157 AA.
AC   O95875;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BAT2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rowen L., Madan A., Qin S., Shaffer T., James R., Ratcliffe A.,
RA   Abbasi N., Dickhoff R., Loretz C., Madan A., Dors M., Young J.,
RA   Lasky S., Hood L.;
RT   "Sequence of the human major histocompatibility complex class III
RT   region.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF129756; AAD18086.1; -.
SQ   SEQUENCE   2157 AA;  228869 MW;  21B817F5B699B0DB CRC64;

  Query Match             5.0%;  Score 190;  DB 4;  Length 2157;
  Best Local Similarity   22.7%;  Pred. No. 6.6e-05;
  Matches  125;  Conservative   34;  Mismatches  180;  Indels  212;  Gaps   24;

Qy       3 ERES---GGLGGGAASPPAASP-----FLGLHIA---------SPPNFRLTHDISLEEFE 45
             |||:   |  |  : :  ||::|        : :|          ||       | : ||:|
Db    1263 ERENAARGSEGKPSLTLPASAPAPEEALTTVTVAPAPRRAAAKSPDLSNQNSDQANEEWE 1322

Qy      46 D-EDLSEITDECGISLQCKDTLSLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTP 104
             : |: | |     :    : | |   |    |||||   : | | : ||     |
Db    1323 TASESSDFTSERRGDKEAPPPVLLTPKAVGTPGGGGGGAVPGISA-MSRGDLSQRAKDLS 1381

Qy     105 GAEDDEEDDDEERAARRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTL 164
               :    ||  |||| |     ||   :| | |   |   ||| ||
Db    1382 KRSFSSQRPGMERQNRRPGPGGKAGSSGS--SSGGGGGGPGGRTGPGRGD---------- 1429

Qy     165 NLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPA 224
                                       | ||    :||  |    |: ||         |||   |
Db    1430 -------------------KRSWPSPKNRSRPP----EERPP------GLPLPP---PP 1456

Qy     225 PTTDRGTSTDSPCRRSTA-------------TQMAPPG------GPPAAPPGGRGHSHR 264
           |::      |     : |                :  || |      ||| || |
Db    1457 PSSSAVFRLDQVIHSNPAGIQQALAQLSSRQGSVTAPGGHPRHKPGPPQAPQG------- 1509

Qy     265 DRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAFLPPTESRMSVSSDP----------- 313
                           |  |||   ||             :||||
Db    1510 -------------------PSPRPPTRYEPQRV-------NSGLSSDPHFEEPGPMVRG 1542

Qy     314 ------DPAAY----PSTAGRP--HPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEPPP 361
                 | |       |    ||   | : :||      | |     |    |: ||  | |
Db    1543 VGGTPRDSAGVSPFPPKRRERPPRKPELLQEE------SLP---PPHSSGFLGSKPEGPG 1593

Qy     362 PPRASLSSDTSALS---YDSVKYTLVVDEHAQLELVSLRP--------------CFGDY 403
           |   |  : | ||:   :: :        |      | ||                 |
Db    1594 PQAESRDTGTEALTPHIWNRL--------HTATSRKSYRPSSMEPWMEPLSPFEDVAGTE 1645

Qy     404 SDESDSATVYDNCASVSS-------------------------------------PYESAIGEE 430
             :|||       : |||                                      | |   |
Db    1646 MSQSDSGVDLSGDSQVSSGPCSQRSSPDGGLKGAAEGPPKRPGGSSPLNAVPCEGPPG-- 1703

Qy     431 YEEAPRPQPPA 441
             | ||  |||
Db    1704 -SEPPRRPPPA 1713


RESULT 11
Q96QC6
ID   Q96QC6      PRELIMINARY;      PRT;  2157 AA.
AC   Q96QC6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BAT2 protein.
GN   BAT2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hirakawa M., Yamaguchi H., Imai K., Shimada J.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Shiina S., Tamiya G., Oka A., Inoko H.;
RT   "Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP000505; BAB63391.1; -.
SQ   SEQUENCE   2157 AA;  228858 MW;  C500E03BE15382C9 CRC64;

  Query Match             5.0%;  Score 190;  DB 4;  Length 2157;
  Best Local Similarity   22.7%;  Pred. No. 6.6e-05;
  Matches  125;  Conservative   34;  Mismatches  180;  Indels  212;  Gaps   24;

Qy       3 ERES---GGLGGGAASPPAASP-----FLGLHIA---------SPPNFRLTHDISLEEFE 45
             |||:   |  |  : :  ||::|        : :|          ||       | : ||:|
Db    1263 ERENAARGSEGKPSLTLPASAPGPEEALTTVTVAPAPRRAAAKSPDLSNQNSDQANEEWE 1322

Qy      46 D-EDLSEITDECGISLQCKDTLSLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTP 104
             : |: | |     :    : | |   |    |||||   : | | : ||     |
Db    1323 TASESSDFTSERRGDKEAPPPVLLTPKAVGTPGGGGGGAVPGISA-MSRGDLSQRAKDLS 1381

Qy     105 GAEDDEEDDDEERAARRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTL 164
               :    ||  |||| |     ||   :| | |   |   ||| ||
Db    1382 KRSFSSQRPGMERQNRRPGPGGKAGSSGS--SSGGGGGGPGGRTGPGRGD---------- 1429

Qy     165 NLFPQVPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPA 224
                                       | ||    :||  |    |: ||         |||   |
Db    1430 -------------------KRSWPSPKNRSRPP----EERPP------GLPLPP---PP 1456
```

```
Qy     225 PTTDRGTSTDSPCRRSTA--------------TQMAPPG------GPPAAPPGGRGHSHR 264
           |::      |     : |              :  || |      ||| || |
Db    1457 PSSSAVFRLDQVIHSNPAGIQQALAQLSSRQGSVTAPGGHPRHKPGPPQAPQG------- 1509

Qy     265 DRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAFLPPTESRMSVSSDP----------- 313
                           |  |||   ||            :||||
Db    1510 --------------------PSPRPPTRYEPQRV-------NSGLSSDPHFEEPGPMVRG 1542

Qy     314 ------DPAAY----PSTAGRP--HPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEPPP 361
                |  |      |    ||   | : :||     | |    |   |: || |  |
Db    1543 VGGTPRDSAGVSPFPPKRRERPPRKPELLQEE------SLP---PPHSSGFLGSKPEGPG 1593

Qy     362 PPRASLSSDTSALS---YDSVKYTLVVDEHAQLELVSLRP--------------CFGDY 403
           |   |  : | ||:   :: :        |      | ||                |
Db    1594 PQAESRDTGTEALTPHIWNRL--------HTATSRKSYRPSSMEPWMEPLSPFEDVAGTE 1645

Qy     404 SDESDSATVYDNCASVSS--------------------------------PYESAIGEE 430
             :|||       : |||                              | |   |
Db    1646 MSQSDSGVDLSGDSQVSSGPCSQRSSPDGGLKGAAEGPPKRPGGSSPLNAVPCEGPPG-- 1703

Qy     431 YEEAPRPQPPA 441
             | ||  |||
Db    1704 -SEPPRRPPPA 1713
```

RESULT 12
Q9GLM4
ID Q9GLM4 PRELIMINARY; PRT; 1715 AA.
AC Q9GLM4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tensin.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20480062; PubMed=11023826;
RA Chen H., Ishii A., Wong W.K., Chen L.B., Lo S.H.;
RT "Molecular characterization of human tensin.";
RL Biochem. J. 351:403-411(2000).
DR EMBL; AF225897; AAG33701.1; -.
DR HSSP; P00519; 1AB2.
DR InterPro; IPR000340; DS_phosphatase.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR000980; SH2.
DR Pfam; PF00017; SH2; 1.
DR ProDom; PD000093; SH2; 1.
DR SMART; SM00462; PTB; 1.
DR SMART; SM00012; PTPc_DSPc; 1.
DR SMART; SM00252; SH2; 1.
DR PROSITE; PS50001; SH2; 1.
SQ SEQUENCE 1715 AA; 183507 MW; D32F16D64CEAF610 CRC64;

Query Match 4.9%; Score 184; DB 6; Length 1715;
Best Local Similarity 20.6%; Pred. No. 0.00013;
Matches 177; Conservative 76; Mismatches 329; Indels 276; Gaps 35;

```
Qy      12 GAASP-PAASPFLGLHIASPPNFR----LTHDISLEEFEDEDLSEITDECGISLQCKDTL 66
           |  ||    || | | ||  |  |      |    | ||              : : |
Db     920 GVRSPVQCVSPELALTIALNPGGRPKEPHLHSYK-EAFE--------------EMEGTS 963

Qy      67 SLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDD-------EEDDDEERAA 119
              || :|: |  |           |    : | |  |     |       |     |
Db     964 PTSPPPSGVRSPPG----------LAKTPLSALGLKPHNPADILLHPTGEPRSYVESVV 1012

Qy     120 RRPGAGP----PKAESGQEPASRGQGQSQ--------GQSQGPGSGDTYRP---KRPTTL 164
           |   |||    |: :|   ||::  |           |    |    |  |      |::
Db    1013 RTAVAGPRTQEPEPKSFSAPAAQAYGHETPLRIGTLGGSFVSPSPLSTSSPILSADSTSV 1072

Qy     165 NLFPQ------------VPRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQ-------- 204
             ||              |       : |||:      |  :|||||  |
Db    1073 GSFPSGESSDQGARTPTQPLLDSGFRSGSLGQPSPLAQRNYQSSSPLPTAGSSYSSPDYS 1132

Qy     205 -------------------------TPPHEHICLSEELPPQ------------------- 220
                                    :|   |  :    ||
Db    1133 LQQFSSPEGQARSQFSVAGVHTVPGSPQARHRTVGTNTPPSPGFGRRAVNPSLAAPSSPS 1192

Qy     221 ------SGPAPTTDRGTSTDSPCRRSTATQMAP--------------------PGGP--- 251
                 ||  |   | :  ||   :  |  :|                        || |
Db    1193 LSHRQVMGPLGTGFHGNTGSSPQSSAATTPGSPSLGRHPGAQVSNLHGNVVTRPGSPSLG 1252

Qy     252 ----------------------PAAPPGGR----------GHSHRDR-IHYQADVRLEATE 279
                                  | :|  ||          |     || : |
Db    1253 RHPGAHQGTLASNLHSNAVASPGSPSLGRHLGGSGSVVPGSPSLDRHVPYGG-------- 1304

Qy     280 EIYLTPVQRPPDAAEPTSA--FLPPTESRMSVSSDPDPAAYP---STAGRPHPSISEEEE 334
             | ||  | |  :  :||  :  |:     ||    || ||   | |  | |  :   :
Db    1305 --YSTPEDRRPTLSRQSSASGYQAPSTPSFPVS----PAYYPGLSSPATSPSPDSAAFRQ 1358

Qy     335 GFDCLSSPERAEPPGGGWRGSL-------GEPPPPPRASLSSDTSALSYDSVKYTLVVDE 387
           |    : ||:     |   |||       |:    | ||  |  |  |  |  :||
Db    1359 GSPTPALPEKRRMSMGDRAGSLPNYATVNGKVSSSPVASGMSSPSGGSTVSFSHTL---- 1414

Qy     388 HAQLELVSLRPCFGDYSDESDSATVYDNCASVSSPYESAIGEEYEEAPRPQPPACLSEDS 447
                 | |  ||   :|    :  | |    :::      |  | |  | | :|
Db    1415 ----------PDFSKYSMPDNSP---ETRAKVKFVQDTSKYWYKPEISREQAIALL-KDQ 1460

Qy     448 TPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLFSCIINGEEQEQTHRAIFRFV----PRH 503
             |    :  |  |   :       ||   :       :   :: : ||  :  |:    ||
Db    1461 EPGAFIIRDSHSFRGAYGLAMKVSSPPPTI-----MQQNKKGDMTHELVRHFLIETGPRG 1515

Qy     504 EDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFPAYYAIEVTKEPEHMAALAKNSDW 563
             :|:   ::|    | |  | :    :    : | |     : :|:   |     ||
Db    1516 V-KLKGCPNEPNFGSLSALVYQHSIIPLALPCKLVIPNRDPTDESKDSSGPA--NSTSDL 1572

Qy     564 VDQ---FRVKFLGSVQVPYHKGNDVLCAAMQKI----ATTRRLTVHFNPPSSCVLEISVR 616
           : |     | |: || :    |   :  |: :      |     |||        ::| :
Db    1573 LKQGAACNVLFVNSVDMESLTGPQAISKAISETLAADPTPAATIVHF--------KVSAQ 1624

Qy     617 GVKIGVKADDSQEAKGNKCSHFFQ----LKNISFCGYHPKNNKY-----------FGFIT 661
           |: :            |:   ||:    |  ::||   |:  |:           |||:
Db    1625 GITL----------TDNQRKLFFRRHYPLNTVTFCDLDPQERKWTKTEGGAPAKLFGFVA 1674

Qy     662 -KHPADHRFACHVFVSED 678
            |  :    |||:|   |
Db    1675 RKQGSATDNACHLFAELD 1692

RESULT 13
O60593
ID O60593 PRELIMINARY; PRT; 640 AA.
AC O60593;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Arg/Abl-interacting protein ArgBP2b (Fragment).
GN ARGBP2B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97362243; PubMed=9211900;
RA Wang B., Golemis E.A., Kruh G.D.;
RT "ArgBP2, a multiple Src homology 3 domain-containing, Arg/Abl-
RT interacting protein, is phosphorylated in v-Abl-transformed cells and
RT localized in stress fibers and cardiocyte Z-disks.";
RL J. Biol. Chem. 272:17542-17550(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Wang B., Golemis E.A., Kruh G.D.;
```

```
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
DR   EMBL; AF049885; AAC05509.1; -.
DR   HSSP; P29354; 1GRI.
DR   InterPro; IPR000108; Neu_cyt_fact_2.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR003127; Sorb.
DR   Pfam; PF00018; SH3; 2.
DR   Pfam; PF02208; Sorb; 1.
DR   PRINTS; PR00499; P67PHOX.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   ProDom; PD000066; SH3; 2.
DR   ProDom; PD016158; Sorb; 1.
DR   SMART; SM00326; SH3; 2.
DR   SMART; SM00459; Sorb; 1.
DR   PROSITE; PS50002; SH3; 2.
KW   SH3 domain.
FT   NON_TER     640     640
SQ   SEQUENCE   640 AA;  70660 MW;  030A8C5036331674 CRC64;

  Query Match             4.8%;  Score 183.5;  DB 4;  Length 640;
  Best Local Similarity   22.0%;  Pred. No. 4e-05;
  Matches  142;  Conservative   70;  Mismatches  243;  Indels  189;  Gaps   31;

Qy       8 GLGGGAASPPA--ASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISLQCKDT 65
Db      38 GNSGGAVSPMSYYQRPF------SPSAYSLPASLN---------SSIVMQHGTSLDSTDT 82

Qy      66 LSLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEE------RAA 119
Db      83 Y---PQHAQSLD----------------------GTTSSSIPLYRSSEEEKRVTVIKAP 116

Qy     120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
Db     117 HYPGIGPVD-ESGIPTAIR----------------------TT------VDRPKDWYKT 146

Qy     180 --NSLGKKHSWQDRVSRSSSP--LKTGEQTPPHEHICLSEELPPQSGPAPTT------DR 229
Db     147 MFKQIHMVHKPDDDTDMYNTPYTYNAGLYNPPYS---------AQSHPAAKTQTYRPLSK 197

Qy     230 GTSTDSP-CRRSTATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQR 288
Db     198 SHSDNSPNAFKDASSPVPPPHVPPPVPP----LRPRDR---------SSTEKHDWDPPDR 244

Qy     289 PPDA----AEPTSAF-LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPE 343
Db     245 KVDTRKFRSEPRSIFEYEPGKS--SILQHERPPPLPTT-----PTPVPREPGRKPLSSSR 297

Qy     344 RAEPPGGGWRGSLGEPPPPPRASL---SSDTSALS--YDSVKYTLVVDEHAQLELVSLRP 398
Db     298 LGE--------VTGSPSPPPRSGAPTPSSRAPALSPTRPPKKPLDYVQDHS--------- 340

Qy     399 CFGDYSDESDSATVYDNC--ASVSSPYESA-----IGEEYEEAPRPQPPACLSEDS---- 447
Db     341 -----SGVFNEASLYQSSIDRSLERPMSSASMASDFRKRRKSEPAVGPPRGLGDQSASRT 395

Qy     448 TPDEPDV-----HFSKKFLNVFMSGRSRSSSAESFGLFSCIING-------------EEQ 489
Db     396 SPGRVDLPGSSTTLTKSFTSSSPSSPSRAKDRESPRSYSSTLTDMGRSAPRERRGTPEKE 455

Qy     490 EQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFPAYYAIEVT- 548
Db     456 KLPAKAVYDFKAQTSKELSFKKGDTVYILRKIDQNWYEGEH--HGRVGIFPISYVEKLTP 513

Qy     549 ------KEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVL 586
Db     514 PEKAQPARPPPPAQPGEIGEAIAKYNFNADTNVELSLRKGDRVI 557


RESULT 14
Q9XW25
ID   Q9XW25      PRELIMINARY;      PRT;  1634 AA.
AC   Q9XW25;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Y18D10A.1 protein.
GN   Y18D10A.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Harris B.R.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; AL034393; CAA22308.1; -.
DR   InterPro; IPR000637; AT_hook.
DR   SMART; SM00384; AT_hook; 1.
DR   PROSITE; PS00354; HMGI_Y; UNKNOWN_1.
SQ   SEQUENCE   1634 AA;  179059 MW;  DC20372F3AAF48D0 CRC64;

  Query Match             4.8%;  Score 182;  DB 5;  Length 1634;
  Best Local Similarity   21.5%;  Pred. No. 0.00018;
  Matches   98;  Conservative   57;  Mismatches  168;  Indels  132;  Gaps   18;

Qy      72 RAGLLSAGGGGAGS----------------RLQAEMLQMDLIDATGDTPGAEDDEEDDD 114
Db     378 KKALLGAQGAAGASEPGSSSSIHGKKGKTVRELQTEMMKGESV---------------- 420

Qy     115 EERAARRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQ 174
Db     421 KEKAARM------RAE-----ASAGRSQAPGPAPAPAASELQDP--PQDFGLSMSDPGSD 467

Qy     175 DTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTD 234
Db     468 SETEEHQKQEKHI-PAMVTRRSARLSALPVTP--KKASSSSKMPPPPSPSPST------- 517

Qy     235 SPCRRS----TATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPP 290
Db     518 -PGRRGRRPRTLSTMSMEPAAAAVTPAPRG---RPRSRSAAKVS-ENTE-----PLSEAP 567

Qy     291 DAAEPTSAFLPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCL---SSPERAEP 347
Db     568 SAPVKRGRGRPRSRSTMSITEDSEPSTSSTAAKRSKRAESDEEEEQDLKLTNKSPEKPKK 627

Qy     348 PGGGWRGSLGE----------------PPPP----------------------PRAS 366
Db     628 PSKTTEETVGDVLKKRLRDTAKTTATVIHTPGPPLRTRKMERMRAPTAVTSSKKEKPKNA 687

Qy     367 LSSDTSALSYDSVKYTLVVDEHAQLELVSLR-------PCFGDYSDESDSATVYDNCASV 419
Db     688 GSADSSINEEEHEDETMILEEQT-LDLPQQTSQQEPRISCGSELLDEQFDA--------- 737

Qy     420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDV 454
Db     738 SEEHSGTVPSAPELTKNPAPPVPEASEASAEPPKI 772


RESULT 15
Q28224
ID   Q28224      PRELIMINARY;      PRT;  1251 AA.
AC   Q28224;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   IRS-1(COS).
```

```
OS   Cercopithecus aethiops (Green monkey) (Grivet).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercopithecus.
OX   NCBI_TaxID=9534;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96067568; PubMed=7488107;
RA   Wang L., Hayashi H., Mitani Y., Ishii K., Ohnishi T., Niwa Y.,
RA   Kido H., Ebina Y.;
RT   "Cloning of a cDNA encoding a 190-kDa insulin receptor substrate-1-
RT   like protein of simian COS cells.";
RL   Biochem. Biophys. Res. Commun. 216:321-328(1995).
DR   EMBL; D64157; BAA11026.1; -.
DR   HSSP; P35568; 1IRS.
DR   InterPro; IPR002404; Insulin_receptS1.
DR   InterPro; IPR001849; PH.
DR   Pfam; PF02174; IRS; 1.
DR   Pfam; PF00169; PH; 1.
DR   PRINTS; PR00628; INSULINRSI.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00310; PTBI; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
SQ   SEQUENCE   1251 AA;  133054 MW;  924CCAC3BE68EB98 CRC64;

  Query Match             4.8%;  Score 180.5;  DB 6;  Length 1251;
  Best Local Similarity   24.4%;  Pred. No. 0.00016;
  Matches  126;  Conservative   56;  Mismatches  196;  Indels  139;  Gaps   28;

Qy      53 TDECGISLQCKDTLSLRPPRAGLLSAGGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEED 112
           | :|    :   ::|  |   | :|:   |:          :    |: | || |  :||
Db     403 TSDCLFPRRSSASVSGSPSDGGFISSDEYGSSPCDFRSSFRSVTPDSLGHTPPARGEEE- 461

Qy     113 DDEERAARRPGAGPP--KAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQV 170
                 | ||    | :|    |||           |:|  | |   | |   | :
Db     462 --LSNYICMGGKGPSTLTAPNGHYILSRG----------GNGHRYTP--GTGLGTSPAL 506

Qy     171 ----PRSQDTLNNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPT 226
               |   |:|    : |      |  :||  | ::||    :   ||
Db     507 AGDEASSAADLDNRFRKRTH------SAGTSPTITHQKTPSQSSVASIEEY--------- 551

Qy     227 TDRGTSTDSPCRRSTATQMAPPGGPPAAPPGGR--GHSHRDRI--HYQADVRLEATEEIY 282
                      |:|  |  ||    |||  || |   :  |  :  ||
Db     552 ----------------TEMM-PAYPPGGGSGGRLPGHRHSAFVPTHSYPEEGLE------ 588

Qy     283 LTPVQRPPDAAEPTSAFLPPTESRMSVSSDPDPAAYPST---AGRPHP----SISEEEEG 335
           : |::|      | |: |   :  | :|  |  |  ||:   :|   |    |:|  ::
Db     589 MHPLERRGGHHRPDSSTLHTDDGYMPMS--PGVAPVPSSRKGSGDYMPMSPKSVSAPQQI 646

Qy     336 FDCL-SSPERAEP-------PGGGWRGSLGEPPPPPRASLSSDT--SALSYDSVKYTLVV 385
            : :   |:| :|       | ||    :|  |    :| || |  |  ||  : :|  |
Db     647 INPIRRHPQRVDPNGYMMMSPSGGCSPDIGGGP----SSSSSSTVPSGSSYGKL-WTKGV 701

Qy     386 DEHAQLELV-----------SLRPCFGDYSDES---DSATVYDNCASVSSPYESAIGEEY 431
             |    |:            | || ||| : |   || |        ||| :   |
Db     702 GAHNSQVLLHPKPPVESSGGKLLPCTGDYMNMSPVGDSNT--------SSPSDCYYG--- 750

Qy     432 EEAPRPQP-------PACLSEDSTPDEPDVHFSKKFLNV-FMSGR----------SRSSS 473
            | |: :|       |       | ||:     : | :   |||          | |:|
Db     751 PEDPQHKPVLSYYSLPRSFKHTQRPGEPEEGARHQHLRLSTSSGRLLYAATADDSSSSTS 810

Qy     474 AESFGLFSCIINGEEQEQT-----HRAIFRFVPRHED 505
           ::| |   |   |   | :     |: :   :||  |
Db     811 SDSLGGGYC---GARLEPSLPHPHHQVLQPHLPRKVD 844


Search completed: December 24, 2002, 08:53:18
Job time : 115 secs
```

OM protein - protein search, using sw model

Run on: December 24, 2002, 08:48:50 ; Search time 18 Seconds
(without alignments)
1162.205 Million cell updates/sec

Title: US-09-966-561-2
Perfect score: 3792
Sequence: 1 MAERESGGLGGGAASPPAAS.....QQFYKQFVEYTCPTEDIYLE 711

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3503	92.4	714	2	US-08-859-201-2	Sequence 2, Appli
2	3263.5	86.1	660	3	US-08-819-177-1	Sequence 1, Appli
3	3245	85.6	659	4	US-09-562-737-11	Sequence 11, Appl
4	2875	75.8	659	4	US-09-562-737-20	Sequence 20, Appl
5	2864	75.5	659	4	US-09-562-737-16	Sequence 16, Appl
6	2846	75.1	659	4	US-09-562-737-13	Sequence 13, Appl
7	2845	75.0	659	4	US-09-562-737-17	Sequence 17, Appl
8	2829	74.6	659	4	US-09-562-737-12	Sequence 12, Appl
9	2824	74.5	659	4	US-09-562-737-14	Sequence 14, Appl
10	2817	74.3	659	4	US-09-562-737-15	Sequence 15, Appl
11	2816	74.3	659	4	US-09-562-737-18	Sequence 18, Appl
12	2814	74.2	659	4	US-09-562-737-19	Sequence 19, Appl
13	1655	43.6	314	2	US-08-859-201-8	Sequence 8, Appli
14	1427	37.6	270	2	US-08-859-201-4	Sequence 4, Appli
15	1279.5	33.7	830	4	US-09-562-737-31	Sequence 31, Appl
16	1116.5	29.4	830	4	US-09-562-737-36	Sequence 36, Appl
17	1111.5	29.3	830	4	US-09-562-737-33	Sequence 33, Appl
18	1103.5	29.1	829	4	US-09-562-737-40	Sequence 40, Appl
19	1097.5	28.9	830	4	US-09-562-737-38	Sequence 38, Appl
20	1093.5	28.8	830	4	US-09-562-737-34	Sequence 34, Appl
21	1083.5	28.6	830	4	US-09-562-737-32	Sequence 32, Appl
22	1075.5	28.4	830	4	US-09-562-737-37	Sequence 37, Appl
23	1074	28.3	830	4	US-09-562-737-39	Sequence 39, Appl
24	1060.5	28.0	830	4	US-09-562-737-35	Sequence 35, Appl
25	751	19.8	155	3	US-08-819-177-4	Sequence 4, Appli
26	392	10.3	79	3	US-08-819-177-16	Sequence 16, Appl
27	389	10.3	77	3	US-08-819-177-17	Sequence 17, Appl
28	359	9.5	76	3	US-08-819-177-15	Sequence 15, Appl
29	327	8.6	68	3	US-08-819-177-18	Sequence 18, Appl
30	305	8.0	59	3	US-08-819-177-19	Sequence 19, Appl
31	261	6.9	49	3	US-08-819-177-20	Sequence 20, Appl
32	260	6.9	50	3	US-08-819-177-13	Sequence 13, Appl
33	207	5.5	39	3	US-08-819-177-21	Sequence 21, Appl
34	186	4.9	696	3	US-08-906-865-4	Sequence 4, Appli
35	186	4.9	696	4	US-09-129-668-4	Sequence 4, Appli
36	175.5	4.6	1243	2	US-08-557-139-2	Sequence 2, Appli
37	169.5	4.5	1185	4	US-09-041-886-23	Sequence 23, Appl
38	169	4.5	1234	2	US-08-317-310A-15	Sequence 15, Appl
39	169	4.5	1234	5	PCT-US95-13041-15	Sequence 15, Appl
40	167.5	4.4	1958	1	US-07-945-283-2	Sequence 2, Appli
41	166.5	4.4	1155	1	US-08-094-948A-29	Sequence 29, Appl
42	166.5	4.4	1155	5	PCT-US96-09319-29	Sequence 29, Appl
43	165.5	4.4	548	2	US-08-469-412A-2	Sequence 2, Appli
44	165.5	4.4	548	4	US-09-021-715-2	Sequence 2, Appli
45	165	4.4	1291	4	US-09-150-460B-10	Sequence 10, Appl

ALIGNMENTS

```
RESULT 1
US-08-859-201-2
; Sequence 2, Application US/08859201
; Patent No. 5880261
;  GENERAL INFORMATION:
;    APPLICANT:  Waeber, G
;    APPLICANT:  Nicod, P
;    TITLE OF INVENTION:  Transcription Factor Islet-Brain 1 (IB1)
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Luann Cserr
;      STREET:  750 Arimo Avenue
;      CITY:  Oakland
;      STATE:  California
;      COUNTRY:  USA
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/859,201
;      FILING DATE:  20-MAY-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9706731.8
;      FILING DATE:  03-APR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9709920.4
;      FILING DATE:  15-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luann Cserr
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  ME A9702
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  714 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-859-201-2

  Query Match             92.4%;  Score 3503;  DB 2;  Length 714;
  Best Local Similarity   93.3%;  Pred. No. 4.9e-294;
  Matches  664;  Conservative   11;  Mismatches   31;  Indels    6;  Gaps    5;

Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
           |||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8 MAERES-GLSGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 66

Qy      61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
```

```
Db     67 QCKDTLSLRPPRAGLLSAGSSGSAGSRLQAEMLQMDLIDAASDTPGAEDDEEDDD-ELAA 125

Qy    120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179

Db    126 QRPGVGPSKAESGQEPASRSQG--QGQGPGTGCGDTYRPKRPTTLNLFPQVPRSQDTLNN 183

Qy    180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239

Db    184 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSDELPPQGSPVPTQDRGTSTDSPCRR 243

Qy    240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299

Db    244 TAATQMAPPSGPPATAPGGRGHSHRDR-SISADVRLEATEEIYLTPVQRPPDPAEPTSTF 302

Qy    300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359

Db    303 LPPTESRMSVSSDPDPAAYSVTAGRPHPSISEEDEGFDCLSSPEQAEPPGGGWRGSLGEP 362

Qy    360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419

Db    363 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASA 422

Qy    420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479

Db    423 SSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 482

Qy    480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539

Db    483 FSCVINGEEHEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 542

Qy    540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599

Db    543 PAYYAIEVTKEPEHMAALAKNSDWIDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 602

Qy    600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659

Db    603 TVHFNPPSSCVLEISVRGVKIGVKADEAQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 662

Qy    660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711

Db    663 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 714


RESULT 2
US-08-819-177-1
; Sequence 1, Application US/08819177
; Patent No. 6043083
;  GENERAL INFORMATION:
;    APPLICANT:  Davis, Roger J.
;    APPLICANT:  Dickens, Martin
;    TITLE OF INVENTION:  INHIBITORS OF THE JNK SIGNAL
;    TITLE OF INVENTION:  TRANSDUCTION PATHWAY AND METHODS OF USE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/819,177
;      FILING DATE:  28 April 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  07917/037001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-542-5070
;      TELEFAX:  617-542-8906
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  660 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  internal
;    FEATURE:
;      OTHER INFORMATION:  JIP-1 protein
US-08-819-177-1

  Query Match             86.1%;  Score 3263.5;  DB 3;  Length 660;
  Best Local Similarity   87.2%;  Pred. No. 2.1e-273;
  Matches  621;  Conservative   12;  Mismatches   26;  Indels   53;  Gaps    4;

Qy      1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60

Db      1 MAERES-GLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 59

Qy     61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119

Db     60 QCKDTLSLRPPRAGLLSAGSSGSAGSRLQAEMLQMDLIDAAGDTPGAEDDEEEEDDELAA 119

Qy    120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179

Db    120 QRPGVGPPKAESNQDPAPR----SQGQGPGTGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 175

Qy    180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239

Db    176 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSDELPPQGSPVPTQDRGTSTDSPCRR 235

Qy    240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299

Db    236 SAATQMAPPSGPPATAPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDPAEPTSTF 295

Qy    300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359

Db    296 MPPTESRMSVSSDPDPAAYSVTAGRPHPSISEEDEGFDCLSSPERAEPPGGGWRGSLGEP 355

Qy    360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419

Db    356 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASA 415

Qy    420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479

Db    416 SSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 475

Qy    480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539

Db    476 FSCVINGEEHEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 535

Qy    540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599

Db    536 PAYYAIEVTKEPEHMAALAKN---------------------------------------- 556

Qy    600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659

Db    557 --------SCVLEISVRGVKIGVKADDALEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 608

Qy    660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711

Db    609 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 660


RESULT 3
US-09-562-737-11
; Sequence 11, Application US/09562737
```

```
; Patent No. 6428967
; GENERAL INFORMATION:
;  APPLICANT:  Herz, Joachim
;  APPLICANT:  Gotthardt, Michael
;  TITLE OF INVENTION:  LDL Receptor Signaling Pathways
;  FILE REFERENCE:  UTSW0708
;  CURRENT APPLICATION NUMBER: US/09/562,737
;  CURRENT FILING DATE:  2000-05-01
;  NUMBER OF SEQ ID NOS:  132
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 11
;   LENGTH: 659
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Sequence
US-09-562-737-11

  Query Match             85.6%;  Score 3245;  DB 4;  Length 659;
  Best Local Similarity   87.1%;  Pred. No. 8.4e-272;
  Matches  620;  Conservative   12;  Mismatches   26;  Indels   54;  Gaps    5;

Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAERES-GLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 59

Qy      61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
           |||||||||||||||||||  | |||||||||||||||| ||||||||||::|:| ||
Db      60 QCKDTLSLRPPRAGLLSAGSSGSAGSRLQAEMLQMDLIDAAGDTPGAEDDEEEEDDELAA 119

Qy     120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
           :||| |||||||| |:|| |    |||| | ||||||||||||||||||||||||||||||
Db     120 QRPGVGPPKAESNQDPAPR----SQGQGPGTGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 175

Qy     180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
           ||||||||||||||||||||||||||||||:|||||  | || ||||||||||||
Db     176 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSDELPPQGSPVPTQDRGTSTDSPCRR 235

Qy     240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
           | |||||| ||||  ||||||||||||||||||||||||||||||||||||| ||||| |
Db     236 SAATQMAPPSGPPATAPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDPAEPTSTF 295

Qy     300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
           :|||||||||||||||||  ||||||||||||:|||||||||||||||||||||||||||
Db     296 MPPTESRMSVSSDPDPAAYSVTAGRPHPSISEEDEGFDCLSSPERAEPPGGGWRGSLGEP 355

Qy     360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     356 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASA 415

Qy     420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
           |||||||||||||||:|:|| ||||||||||||| ||||||||||||||||||||||||||
Db     416 SSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPDV-FSKKFLNVFMSGRSRSSSAESFGL 474

Qy     480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
           |||:|||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     475 FSCVINGEEHEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 534

Qy     540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
           ||||||||||||||||||||
Db     535 PAYYAIEVTKEPEHMAALAKN---------------------------------------- 555

Qy     600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
                    ||||||||||||||||||: ||||||||||||||||||||||||||||||||
Db     556 --------SCVLEISVRGVKIGVKADDALEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 607

Qy     660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     608 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 659


RESULT 4
US-09-562-737-20
; Sequence 20, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
;  APPLICANT:  Herz, Joachim
;  APPLICANT:  Gotthardt, Michael
;  TITLE OF INVENTION:  LDL Receptor Signaling Pathways
;  FILE REFERENCE:  UTSW0708
;  CURRENT APPLICATION NUMBER: US/09/562,737
;  CURRENT FILING DATE:  2000-05-01
;  NUMBER OF SEQ ID NOS:  132
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 20
;   LENGTH: 659
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Sequence
US-09-562-737-20

  Query Match             75.8%;  Score 2875;  DB 4;  Length 659;
  Best Local Similarity   78.8%;  Pred. No. 7.6e-240;
  Matches  561;  Conservative   18;  Mismatches   79;  Indels   54;  Gaps    5;

Qy       1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
           |||||| ||| ||||||||| |||||||| |||||||| |||||||| |||||||||
Db       1 MAERES-GLGAGAASPPAASDFLGLHIASPENFRLTHDISFEEFEDEDLSGITDECGISL 59

Qy      61 QCKDTLSLRPPRAGLLSAGG-GGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
           |||||||||  |||||||  | ||||||| |||||||| |||||||| |::|:| ||
Db      60 HCKDTLSLRPIRAGLLSAGSKGSAGSRLQALMLQMDLIDAMGDTPGAEDDNEEEDDELAA 119

Qy     120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
           |||| ||||| | |:|| |    | ||  | ||| |||||||| |||||||| |||||
Db     120 RRPGVGPPKASSNQDPAPR----STGQGPGTGSGVTYRPKRPTTWNLFPQVPRSYDTLNN 175

Qy     180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
           |||| ||||||||| |||||||||:||||||||  :|||||  | || |||||| |||||
Db     176 NSLGAKHSWQDRVSDSSSPLKTGEETPPHEHICLFDELPPQGSPGPTQDRGTSTHSPCRR 235

Qy     240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
           | || |||| ||||  |||||||| |||||||| |||||||| ||||||| |:||| |
Db     236 SAATIMAPPSGPPAKAPGGRGHSHLDRIHYQADVMLEATEEIYLNPVQRPPDPAQPTSTF 295

Qy     300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
           :||| |||||||| ||||  ||| |||||||||: |||||||||| |||||||| |||||
Db     296 MPPTRSRMSVSSDPSPAAYSVTAGTPHPSISEEDVGFDCLSSPEWAEPPGGGWRYSLGEP 355

Qy     360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
           |||| |||||||||| |||||||| |||||||| |||||||| |||||||| ||||
Db     356 PPPPAASLSSDTSADSYDSVKYTLEVDEHAQLELFSLRPCFGDYGDESDSATVYHNCASA 415

Qy     420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
           |||| |||||||| |:|:|| || |||||||||: |||||||| ||||||||:||||
Db     416 SSPYISAIGEEYEEKPQPRPPTCLLEDSTPDEPDM-FSKKFLNVFNSGRSRSSSAQSFGL 474

Qy     480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
           |||:| ||| ||||:|||||||| |||||||||:|||||||||:|||||||| ||||
Db     475 FSCVIRGEEHEQTHRSIFRFVPRHETELELEVDDPVLVELQAEDYYYEAYNMRTGDRGVF 534

Qy     540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
           ||||| |||||||| |||||
Db     535 PAYYAEEVTKEPEHMFALAKN---------------------------------------- 555

Qy     600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
                    |||| |||||||| |||: ||| |||||||| |||||||| |||||||
Db     556 --------SCVLGISVRGVKIGHKADDALEAKINKCSHFFQLLNISFCGYHPMNNKYFGF 607

Qy     660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
           || |||||||| |||||||| |||||||| |||||||| |||||||||
Db     608 ITNHPADHRFACQVFVSEDSTKRLAESVGRAFSQFYKQFVEYVCPTEDIYLE 659
```

```
RESULT 5
US-09-562-737-16
; Sequence 16, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
;  APPLICANT:  Herz, Joachim
;  APPLICANT:  Gotthardt, Michael
;  TITLE OF INVENTION:  LDL Receptor Signaling Pathways
;  FILE REFERENCE:  UTSW0708
;  CURRENT APPLICATION NUMBER: US/09/562,737
;  CURRENT FILING DATE:  2000-05-01
;  NUMBER OF SEQ ID NOS:  132
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 16
;   LENGTH: 659
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Sequence
US-09-562-737-16

  Query Match             75.5%;  Score 2864;  DB 4;  Length 659;
  Best Local Similarity   78.5%;  Pred. No. 6.7e-239;
  Matches  559;  Conservative   23;  Mismatches   76;  Indels   54;  Gaps    5;

Qy      1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
          |||||: ||||||||| |||||||| |||||||| |||||||| ||||||||| |||
Db      1 MAEREA-GLGGGAASPDAASPFLGLHEASPPNFRLTFDISLEEFEDGDLSEITDECHISL 59

Qy     61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
          |||||| ||||||||| ||  | ||| |||||||||:||| ||||| ||||||::|:: ||
Db     60 QCKDTLILRPPRAGLLKAGSSGSAGSLLQAEMLQMDMIDAAGDTPGNEDDEEEEDDQLAA 119

Qy    120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
          :||| | ||||| |:|: |    |||| | ||||||||| ||||||||| |||||||||
Db    120 QRPGVGRPKAESNQDPSPR----SQGQGPGVGSGDTYRPKWPTTLNLFPQYPRSQDTLNN 175

Qy    180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
          |||||||||||||||||:||||||||| ||||:|||   | || || |||||||||
Db    176 ASLGKKHSWQDRVSRSSSPLETGEQTPPHEFICLSDELPPGGSPVPTQDRHTSTDSPCRR 235

Qy    240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
          |||||||  |||  |||| ||||||||| |||||||| ||||||||| | ||||| |
Db    236 IAATQMAPPSKPPATAPGGRLHSHRDRIHYMADVRLEATENIYLTPVQRPQDPAEPTSTF 295

Qy    300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
          |||||||||:|||||||| |||||||| ||:|||||:|||||||| |||||||||
Db    296 RPPTESRMSVTSDPDPAAYSWTAGRPHPSIYEEDEGFDCLASPERAEPPGDGWRGSLGEP 355

Qy    360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
          ||||||||| ||||||||| |||||||| ||||||||| ||||||||| ||||||||
Db    356 EPPPRASLSSFTSALSYDSVGYTLVVDEHAHLELVSLRPCIGDYSDESDSKTVYDNCASA 415

Qy    420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
          ||||||||| |||||:|:|  |||||||| |||| ||||| |||||||||:|:||||||||
Db    416 LSPYESAIGEMYEEAPQPRPNTCLSEDSTPQEPDV-FSKKFRNVFMSGRSRTSSAESFGL 474

Qy    480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
          | |:|||| | ||||||||| ||||||||| ||||||||| |||||||| ||||||||
Db    475 FVCVINGEEHEWTHRAIFRFVYRHEDELELEADDPLLVELQDEDYWYEAYNERTGARGVF 534

Qy    540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
          | ||||||||| |||||||||
Db    535 PFYYAIEVTKEGEHMAALAKN---------------------------------------- 555

Qy    600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
                |||||||||:||||||||::||||||||| ||||||||| ||||||||| |
Db    556 --------HCVLEISVRGIKIGVKADDAKEAKGNKCSHLFQLKNISFCMYHPKNNKYFNF 607

Qy    660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          |||||||||:|||||||| |||||||| |||||||| |||||||||:|||
Db    608 ITKHPADHQFACHVFVSERSTKALAESVSRAFQQFYKQTVEYTCPTEDVYLE 659


RESULT 6
US-09-562-737-13
; Sequence 13, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
;  APPLICANT:  Herz, Joachim
;  APPLICANT:  Gotthardt, Michael
;  TITLE OF INVENTION:  LDL Receptor Signaling Pathways
;  FILE REFERENCE:  UTSW0708
;  CURRENT APPLICATION NUMBER: US/09/562,737
;  CURRENT FILING DATE:  2000-05-01
;  NUMBER OF SEQ ID NOS:  132
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 659
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Sequence
US-09-562-737-13

  Query Match             75.1%;  Score 2846;  DB 4;  Length 659;
  Best Local Similarity   78.4%;  Pred. No. 2.4e-237;
  Matches  558;  Conservative   20;  Mismatches   80;  Indels   54;  Gaps    5;

Qy      1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
          || ||| |||||| ||||||||| |||||||| |||||||| |||||| |||||||||
Db      1 MAARES-GLGGGADSPPAASPFLELHIASPPNFFLTHDISLEEGEDEDLSEITHECGISL 59

Qy     61 QCKDTLSLRPPRAGLLSAGGGGA-GSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
          ||| ||||||||| ||||  |: ||||||||| |||||| || |||||||: |:| ||
Db     60 QCKITLSLRPPRAKLLSAGSSGSLGSRLQAEMLMMDLIDAAGDNPGAEDDEEEQDDELAA 119

Qy    120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
          :|| |||||||  :|| |    |||| | |||||| |||||||||:||||||||| ||
Db    120 QRPRVGPPKAESNSDPAPR----SQGQTPGTGSGDTYVPKRPTTLNLWPQVPRSQDTYNN 175

Qy    180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
          |||||||:||||||||| ||||||||| |||||||:| |||  | || ||||||||| ||
Db    176 NSLGKKHAWQDRVSRSSDPLKTGEQTPEHEHICLSDEFPPQGSPVPTGDRGTSTDSPHRR 235

Qy    240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
          | ||||| | ||||  | |||||||||:||||||||| ||||||||| |||| ||||  |
Db    236 SAATQMAIPSGPPATAPKGRGHSHRDRLHYQADVRLEMTEEIYLTPVNRPPDPAEPTQTF 295

Qy    300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
          :|||||| |||||||||:|  |||||| |||||:||| |||||||||  |||||||||| ||
Db    296 MPPTESRRSVSSDPDPASYSVTAGRPHTSISEEDEGFVCLSSPERAEWPGGGWRGSLYEP 355

Qy    360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
          ||||||| |||||||||||||||||| ||||||||| ||||||||| |||||||||  |
Db    356 PPPPRASASSDTSALSYDSVKYTLVVDFHAQLELVSLGPCFGDYSDEHDSATVYDNCISA 415

Qy    420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
          ||||||| ||||||||: :|| ||||| ||||||| || |||||||||:||||||||  |
Db    416 SSPYESAKGEEYEEAPQLRPPTCLSEDMTPDEPDV-FSNKFLNVFMSGQSRSSSAESFRL 474

Qy    480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
          |||:|||| |||||||| |||||||| |||||||| |||||||| |||||||| |
Db    475 FSCVINGESHEQTHRAIFTFVPRHEDELVLEVDDPLLVWLQAEDYWYEYYNMRTGARGAF 534

Qy    540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
          |||||||| |||||||| ||
Db    535 PAYYAIEVDKEPEHMAALEKN---------------------------------------- 555

Qy    600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
                |||||| |||||||| |: |||||| |||||||| |||||||| ||||
Db    556 --------SCVLEISFRGVKIGVKAGDALEAKGNKHSHFFQLKNIIFCGYHPKNNLYFGF 607
```

```
Qy    660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          ||||| ||||||||| |||||||||:||||||||| ||||||||||:||||||
Db    608 ITKHPMDHRFACHVFNSEDSTKALAQSVGRAFQQFRKQFVEYTCPSEDIYLE 659


RESULT 7
US-09-562-737-17
; Sequence 17, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
;  APPLICANT:  Herz, Joachim
;  APPLICANT:  Gotthardt, Michael
;  TITLE OF INVENTION:  LDL Receptor Signaling Pathways
;  FILE REFERENCE:  UTSW0708
;  CURRENT APPLICATION NUMBER: US/09/562,737
;  CURRENT FILING DATE:  2000-05-01
;  NUMBER OF SEQ ID NOS:  132
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 17
;   LENGTH: 659
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Sequence
US-09-562-737-17

  Query Match             75.0%;  Score 2845;  DB 4;  Length 659;
  Best Local Similarity   78.6%;  Pred. No. 2.9e-237;
  Matches  561;  Conservative   19;  Mismatches   76;  Indels   58;  Gaps    6;

Qy      1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
          ||||||  ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||
Db      1 MAERES-ALGGGAASPPDASPFLGLHIESPPNFRLTHFISLEEFEDEGLSEITDECGHSL 59

Qy     61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
          |||||||:||||||||| |  | ||||:||||||||| || ||||||:||||::|:||||
Db     60 QCKDTLSIRPPRAGLLSKGSSGSAGSRMQAEMLQMDLNDAAGDTPGAQDDEEEEDDERAA 119

Qy    120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPG--SGDTYRPKRPTTLNLFPQVPRSQDTL 177
          :||| || |||| |:||:| ||      ||||  ||||||||| ||||||||| ||||||
Db    120 QRPGVGPSKAESNQDPATRSQG------QGPGTVSGDTYRPKRWTTLNLFPQVYRSQDTL 173

Qy    178 NNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPC 237
          |||:|||||||| ||||||||| ||||||||| |||:|||||  | || ||| ||||||
Db    174 NNNALGKKHSWQDDVSRSSSPLKEGEQTPPHEHFCLSDELPPQHSPVPTQDRGISTDSPC 233

Qy    238 RRSTATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTS 297
          ||| ||||||| | ||  ||||| ||||||||| ||||||||| |||||||||  |||||
Db    234 RRSKATQMAPPSGLPATAPGGRGMSHRDRIHYQNDVRLEATEEQYLTPVQRPPRPAEPTS 293

Qy    298 AFLPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLG 357
          |: |||||||||:|||||||   ||||||||| |:||||||| ||||||||| ||||||
Db    294 TFMSPTESRMSVSTDPDPAAYSVVAGRPHPSISWEDEGFDCLSYPERAEPPGGAWRGSLG 353

Qy    358 EPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCA 417
          ||| |||||||||| |||||||||:||||||||| |||||||| |||||||| ||||||
Db    354 EPPDPPRASLSSDESALSYDSVKFTLVVDEHAQGELVSLRPCFHDYSDESDSAIVYDNCA 413

Qy    418 SVSSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESF 477
          | | |||||||||| ||||:|:|| |||||||||  ||| |||||| ||||||||| |||||
Db    414 SASKPYESAIGEELEEAPQPRPPMCLSEDSTPDNPDV-FSKKFLQVFMSGRSRSRSAESF 472

Qy    478 GLFSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARG 537
          |||| :||||| || |||||||| |||||||| |||||||| |||||||| |||||
Db    473 GLFSSVINGEEHEQVHRAIFRFVPWHEDELELEVYDPLLVELQAADYWYEAYNMDTGARG 532

Qy    538 VFPAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTR 597
          |||| |||||||| |||||||||
Db    533 VFPAEYAIEVTKEPFHMAALAKNSG---------------------------------- 557

Qy    598 RLTVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYF 657
                    ||||||||| |||||||:  ||||||||| ||||||||| ||||||||
Db    558 ------------VLEISVRGVHIGVKADDALIAKGNKCSHFKQLKNISFCGLHPKNNKYF 605

Qy    658 GFITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          | ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||
Db    606 GMITKHPADHRNACHVFVSEDQTKALAESVGSAFQQFYKQFTEYTCPTEDIVLE 659


RESULT 8
US-09-562-737-12
; Sequence 12, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
;  APPLICANT:  Herz, Joachim
;  APPLICANT:  Gotthardt, Michael
;  TITLE OF INVENTION:  LDL Receptor Signaling Pathways
;  FILE REFERENCE:  UTSW0708
;  CURRENT APPLICATION NUMBER: US/09/562,737
;  CURRENT FILING DATE:  2000-05-01
;  NUMBER OF SEQ ID NOS:  132
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 12
;   LENGTH: 659
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Sequence
US-09-562-737-12

  Query Match             74.6%;  Score 2829;  DB 4;  Length 659;
  Best Local Similarity   78.4%;  Pred. No. 7.1e-236;
  Matches  558;  Conservative   15;  Mismatches   85;  Indels   54;  Gaps    5;

Qy      1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
          | |||| ||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||
Db      1 MDERES-GLGGGEASPPAASPFFGLHIASPPNGRLTHDISLEHFEDEDLSEIIDECGISL 59

Qy     61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
          || ||||||||| |||||  | ||||||||| ||||||| | ||||||||| :|:| ||
Db     60 QCLDTLSLRPPRMGLLSAGSSGNAGSRLQAEMQQMDLIDAAGRTPGAEDDEESEDDELAA 119

Qy    120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
          :| | ||||||| |:|| |    |||   | ||||| ||||||||| ||||||||| |||
Db    120 QRTGVGPPKAESVQDPAPR----SQGWGPGTGSGDTARPKRPTTLNDFPQVPRSQDELNN 175

Qy    180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
          |||||| ||||||||| ||||||||| ||||||||: |||| | | |||||||| |||
Db    176 NSLGKKFSWQDRVSRSGSPLKTGEQTHPHEHICLSDILPPQGSPVPKQDRGTSTDSLCRR 235

Qy    240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
          | |||| || ||||   |||||||||:||||||||| ||||||||| ||||| ||| | |
Db    236 SAATQMMPPSGPPATANGGRGHSHRDQIHYQADVRLRATEEIYLTPSQRPPDPAEPVSTF 295

Qy    300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
          :||||| ||||||||| ||  ||||| |||||:|| |||||||||  ||||||||| |||
Db    296 MPPTESWMSVSSDPDPYAYSVTAGRPAPSISEEDEGDDCLSSPERAFPPGGGWRGSGGEP 355

Qy    360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
          |||||| |||||||||| |||||||||| ||||||||:|||||||||  |||||||||  ||
Db    356 PPPPRAHLSSDTSALSIDSVKYTLVVKEHAQLELVSMRPCFGDYSDNSDSATVYDNQASA 415

Qy    420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
          |||||| |||||||||  |:|| |||| |||||||| | |||||||||  |||||||:||
Db    416 SSPYESRIGEEYEEAPSPRPPTCLSETSTPDEPDV-FVKKFLNVFMSWRSRSSSAESYGL 474

Qy    480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
          |||:||| | ||||||| |||||||| |||||||| |||||||| |||||||| ||
Db    475 FSCVINGAEHEQTHRAIDRFVPRHEDEEELEVDDPLLFELQAEDYWYGAYNMRTGARHVF 534

Qy    540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
          |||||||:||||||||| |||
Db    535 PAYYAIEITKEPEHMAAKAKN---------------------------------------- 555
```

```
Qy    600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
                  |||||| |||||||||| ||: ||||| |||||||||| |||||||||| |||||
Db    556 --------SCVLEILVRGVKIGVKMDDALEAKGNNCSHFFQLKNQSFCGYHPKNRKYFGF 607

Qy    660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          |||| |||||||| |||||||| |||||||||:|||||||| |||||||
Db    608 ITKHSADHRFACHVTVSEDSTKALVESVGRAFQQWYKQFVEYTCYTEDIYLE 659


RESULT 9
US-09-562-737-14
; Sequence 14, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
;  APPLICANT:  Herz, Joachim
;  APPLICANT:  Gotthardt, Michael
;  TITLE OF INVENTION:  LDL Receptor Signaling Pathways
;  FILE REFERENCE:  UTSW0708
;  CURRENT APPLICATION NUMBER: US/09/562,737
;  CURRENT FILING DATE:  2000-05-01
;  NUMBER OF SEQ ID NOS:  132
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 14
;   LENGTH: 659
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Sequence
US-09-562-737-14

  Query Match             74.5%;  Score 2824;  DB 4;  Length 659;
  Best Local Similarity   78.7%;  Pred. No. 1.9e-235;
  Matches  560;  Conservative   20;  Mismatches   78;  Indels   54;  Gaps    5;

Qy      1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
          ||| || ||||||| |||||||| |||||||| |||||||| |||||||| |||||
Db      1 MAEAES-GLGGGAADPPAASPFLGEHIASPPNFRFTHDISLEEFGDEDLSEITDHCGISL 59

Qy     61 QCKDTLSLRPPRAGLLSAGGGGAG-SRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
          |||| |||||||| |||  |:  |||||||| ||||| ||| |||||||:: :| ||
Db     60 QCKDILSLRPPRAGKLSAGSSGSALSRLQAEMLQNDLIDAAGDTQGAEDDEEEERDELAA 119

Qy    120 RRPGAGPPKAESGQEPASRGQGQSQGQSGQPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
          :|||:||||||| | || |    |||| | ||||||| |||||||| |||||||| |
Db    120 QRPGSGPPKAESNQTPAPR----SQGQGVGTGSGDTYRWKRPTTLNLFYQVPRSQDTLAN 175

Qy    180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
          |||||||| |||||||| |||||||| ||||||:|| ||  | ||  |||||||||  |
Db    176 NSLGKKHSDQDRVSRSSSELKTGEQTPPFEHICLSDELGPQGSPVPTQHRGTSTDSPCIR 235

Qy    240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
          | |||||  |||  || |||||||| |||||||| |||||||||:||| ||||| |
Db    236 SAATQMAPKSGPPATAPGLRGHSHRDRIMYQADVRLEANEEIYLTPVQQPPDPAEPTSRF 295

Qy    300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
          :|||||||:|||||||||:  ||||||| ||||:|||| |||||||| |||||||| |
Db    296 MPPTESRMTVSSDPDPAAWSVTAGRPHPYISEEDEGFDALSSPERAEPDGGGWRGSLGFP 355

Qy    360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
          |||||||| |||||||| |||||||| |||||||| |||||||| |||||||| |||||||
Db    356 PPPPRASLGSDTSALSYDHVKYTLVVDEIAQLELVSLRKCFGDYSDESLSATVYDNCAMA 415

Qy    420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
          |||||||| |||||||:|||| |||||| |||||| ||| |||||||||:|||||||||:
Db    416 SSPYESAINEEYEEAPQPQPPTCLSEDSRPDEPDV-FSKSFLNVFMSGRTRSSSAESFGV 474

Qy    480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
          |||:||||| |||||||||:|||||||| |||||||| |||||||| |||||||| |||||||
Db    475 FSCVINGEEWEQTHRAIFRYVPRHEDELEAEVDDPLLVEDQAEDYWYEAENMRTGARGVG 534

Qy    540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
          |||||||||  |||||||||  |
```

```
Db    535 PAYYAIEVTHEPEHMAALAIN---------------------------------------- 555

Qy    600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
                  ||||||||:||||||||| : ||||||| |||||||| |||||||| |||
Db    556 --------SCVLEISVKGVKIGVKADLALEAKGNKCMHFFQLKNISNCGYHPKNNKQFGF 607

Qy    660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          |||||| |||||||||:|||||||| |||||||| |||||||| |||||
Db    608 ITKHPARHRFACHVFVTEDSTKALAEVVGRAFQQFYWQFVEYTCPTYDIYLE 659


RESULT 10
US-09-562-737-15
; Sequence 15, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
;  APPLICANT:  Herz, Joachim
;  APPLICANT:  Gotthardt, Michael
;  TITLE OF INVENTION:  LDL Receptor Signaling Pathways
;  FILE REFERENCE:  UTSW0708
;  CURRENT APPLICATION NUMBER: US/09/562,737
;  CURRENT FILING DATE:  2000-05-01
;  NUMBER OF SEQ ID NOS:  132
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 15
;   LENGTH: 659
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Sequence
US-09-562-737-15

  Query Match             74.3%;  Score 2817;  DB 4;  Length 659;
  Best Local Similarity   78.4%;  Pred. No. 7.7e-235;
  Matches  560;  Conservative   18;  Mismatches   78;  Indels   58;  Gaps    6;

Qy      1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
          |||| | |||||||| |||||||| |||||||| |||||||| |||||||| ||||
Db      1 MAERAS-GLGGGAASDPAASPFLGLEIASPPNFRLFHDISLEEFEGEDLSEITDEHGISL 59

Qy     61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
          |||||:|||||||| ||| | || |||||||| |||| |||| |||||||::|:| ||
Db     60 QCKDTISLRPPRAGLKSAGSSGSAGLRLQAEMLQMMLIDAAGDTPNAEDDEEEEDQELAA 119

Qy    120 RRPGAGPPKAESGQEPASRGQGQSQGQSGQP--GSGDTYRPKRPTTLNLFPQVPRSQDTL 177
          :|||  |||||| |: | | ||      |||  |||||||| |||||||| ||||||||
Db    120 QRPGVRPPKAESNQDSAPRSQG------QGPTTGSGDTYRPVRPTTLNLFPWVPRSQDTL 173

Qy    178 NNNSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPC 237
          | |||||||| |||||||| |||||||| |||||:||| |  | || | ||||||||
Db    174 NYNSLGKKHSWADRVSRSSSPDKTGEQTPPHFHICLSDELPGQGSPVPTQDHGTSTDSPC 233

Qy    238 RRSTATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTS 297
          | | ||||||| |||  ||| ||||||||| |||||||| |||||||| || |||||
Db    234 RISAATQMAPPKGPPATAPGGLGHSHRDRIHMQADVRLEATNEIYLTPVQRQPDPAEPTS 293

Qy    298 AFLPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLG 357
            :|||||||| |||||||||  ||||||||:|||:||||| |||||||| ||||||||
Db    294 TRMPPTESRMSSSSDPDPAAYTVTAGRPHPSVSEEDEGFDCWSSPERAEPPYGGWRGSLG 353

Qy    358 EPPPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCA 417
          | |||||||| |||||||| |||||||| |||||||| |||||||| |||||||| |||||||
Db    354 EAPPPPRASLSDDTSALSYDSEKYTLVVDEHFQLELVSLRPGFGDYSDESDHATVYDNCA 413

Qy    418 SVSSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESF 477
          |:|||||||||:||||||:|: | ||||||| ||||| |||| |||||||||:|||||||
Db    414 SISSPYESAIGKEYEEAPQPRLPTCLSEDSTMDEPDV-FSKKNLNVFMSGRSQSSSAESF 472

Qy    478 GLFSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARG 537
          || ||:|||||  |||||||| |||||||| |||||||| |||||||| |||||||| |||||||
Db    473 GLRSCVINGEEHSQTHRAIFRFTPRHEDELELVVDDPLLVELWAEDYWYEAYYMRTGARG 532
```

```
Qy    538 VFPAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTR 597
          || |||||||||:|||||||||
Db    533 VFAAYYAIEVTKDPEHMAALAK---------------------------------------- 554

Qy    598 RLTVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYF 657
                  ||||||||| ||||||||| ||||||| ||||||||| |||||||||
Db    555 ---------ESCVLEISVRFVKIGVKADDGLEAKGNKCSIFFQLKNISFKGYHPKNNKYL 605

Qy    658 GFITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          ||||||||| ||||||||| ||||||||| |||||||||:||||||||| ||||
Db    606 GFITKHPADMRFACHVFVSNDSTKALAESQGRAFQQFYKRFVEYTCPTESIYLE 659


RESULT 11
US-09-562-737-18
; Sequence 18, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
;  APPLICANT:  Herz, Joachim
;  APPLICANT:  Gotthardt, Michael
;  TITLE OF INVENTION:  LDL Receptor Signaling Pathways
;  FILE REFERENCE:  UTSW0708
;  CURRENT APPLICATION NUMBER: US/09/562,737
;  CURRENT FILING DATE:  2000-05-01
;  NUMBER OF SEQ ID NOS:  132
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 18
;   LENGTH: 659
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Sequence
US-09-562-737-18

  Query Match             74.3%;  Score 2816;  DB 4;  Length 659;
  Best Local Similarity   78.2%;  Pred. No. 9.4e-235;
  Matches  557;  Conservative   18;  Mismatches   83;  Indels   54;  Gaps    5;

Qy      1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
          |||||||  |||||||| |||||||| |||||||| |||||||| |||||||| |
Db      1 MAERESGA-GGGAASPPADSPFLGLHIAEPPNFRLTHDFSLEEFEDEDGSEITDECGIHL 59

Qy     61 QCKDTLSLRPPRAGLLSA-GGGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
          |||||||| |||||||||   | ||||| ||||||||| | |||||||:|||::|:|  |
Db     60 QCKDTLSLIPPRAGLLSAKSSGSAGSRLLAEMLQMDLIMAAGDTPGAENDEEEEDDELQA 119

Qy    120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
          :||| |||:||| |:||      |||| | |:||||||||| ||||||||| |||||||
Db    120 QRPGVGPPRAESNQDPAP----SSQGQGPGTGTGDTYRPKRPVTLNLFPQVPWSQDTLNN 175

Qy    180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
          || |||||||| |||||||| |||||||| ||:|||||  | || |||| |||||||
Db    176 NSYGKKHSWQDRASRSSSPLKTDEQTPPHEHIELSDELPPQGFPVPTQDRGTGTDSPCRR 235

Qy    240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
          |  |||||| || |  |||||| ||||||||| ||||||||| ||||||||| ||||| |
Db    236 SAHTQMAPPSGPIATAPGGRGHKHRDRIHYQALVRLEATEEIMLTPVQRPPDNAEPTSTF 295

Qy    300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
          :| ||||||||| |||||  |:||||||||| :|||||||| |||||||||:|||||||
Db    296 MPQTESRMSVSSRPDPAAYSVTSGRPHPSISETDEGFDCLSSVERAEPPGGGYRGSLGEP 355

Qy    360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
          || |||||||| |||||||| |||||||| ||||||||| ||||||||| ||||||
Db    356 PPAPRASLSSDTDALSYDSVKYELVVDEHAQLFLVSLRPCFGGYSDESDSATHYDNCASA 415

Qy    420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
          || |||||||||:|||:|:||  ||||||||| || ||||||| ||||||||| ||||||
Db    416 SSIYESAIGEEYKEAPQPRPPTLLSEDSTPDEMDV-FSKKFLNNFMSGRSRSSQAESFGL 474

Qy    480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
          ||| ||||| ||| |||||||||| |||||||||| |||||||||| |||||||||| ||||||

Db    475 FSCRINGEEHEQTSRAIFRFVPRTEDELELEVDVPLLVELQAEWYWYEAYNMRYGARGVF 534

Qy    540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
          ||| |||||||||| |||||||
Db    535 PAYAAIEVTKEPEDMAALAKN---------------------------------------- 555

Qy    600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
                  || ||||||||| |||||| | ||||||||| ||||||||| |||||||||
Db    556 --------SCELEISVRGVKFGVKADDALEGKGNKCSHFFHLKNISFCGYIPKNNKYFGF 607

Qy    660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
           ||||||||| ||||||||| ||||||||| |||||||||:||||||||| |
Db    608 KTKHPADHRFLCHVFVSEDSMKALAESVGRNFQQFYKQFVQYTCPTEDIYRE 659


RESULT 12
US-09-562-737-19
; Sequence 19, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
;  APPLICANT:  Herz, Joachim
;  APPLICANT:  Gotthardt, Michael
;  TITLE OF INVENTION:  LDL Receptor Signaling Pathways
;  FILE REFERENCE:  UTSW0708
;  CURRENT APPLICATION NUMBER: US/09/562,737
;  CURRENT FILING DATE:  2000-05-01
;  NUMBER OF SEQ ID NOS:  132
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 19
;   LENGTH: 659
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Sequence
US-09-562-737-19

  Query Match             74.2%;  Score 2814;  DB 4;  Length 659;
  Best Local Similarity   78.2%;  Pred. No. 1.4e-234;
  Matches  556;  Conservative   18;  Mismatches   83;  Indels   54;  Gaps    5;

Qy      1 MAERESGGLGGGAASPPAASPFLGLHIASPPNFRLTHDISLEEFEDEDLSEITDECGISL 60
          |||||| || |||||||| |||||||| |||||||| |||||||| |||||||| |||||||
Db      1 MAERES-GLAGGAASPPAADPFLGLHIASEPNFRLTHDIFLEEFEDEDLGEITDECGISH 59

Qy     61 QCKDTLSLRPPRAGLLSAG-GGGAGSRLQAEMLQMDLIDATGDTPGAEDDEEDDDEERAA 119
          |||||||| |||||||||  | |||||| |||||||| ||||||||:||::|:| |
Db     60 QCKDTLSLRIPRAGLLSAGKSGSAGSRLQLEMLQMDLIDMAGDTPGAEDNEEEEDDELAQ 119

Qy    120 RRPGAGPPKAESGQEPASRGQGQSQGQSQGPGSGDTYRPKRPTTLNLFPQVPRSQDTLNN 179
          :||| |||| || |:|| |    :|||  | || |||||||| |||||||| ||||||
Db    120 QRPGVGPPKRESNQDPAPR----TQGQGPGTGSVDTYRPKRPTWLNLFPQVPRYQDTLNN 175

Qy    180 NSLGKKHSWQDRVSRSSSPLKTGEQTPPHEHICLSEELPPQSGPAPTTDRGTSTDSPCRR 239
          ||| ||||||||| |||||||| |||||||| |:|||||    || ||||| ||||||
Db    176 NSLAKKHSWQDRVDRSSSPLKTGFQTPPHEHICGSDELPPQGSHVPTQDRGTSIDSPCRR 235

Qy    240 STATQMAPPGGPPAAPPGGRGHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEPTSAF 299
          | | ||||| |||   ||||||| |||||||| |||||||| |||||||  |||| |
Db    236 SAAKQMAPPSGPPLTAPGGRGHSMRDRIHYQADNRLEATEEIYQTPVQRPPDPREPTSTF 295

Qy    300 LPPTESRMSVSSDPDPAAYPSTAGRPHPSISEEEEGFDCLSSPERAEPPGGGWRGSLGEP 359
          :||:||||||||| ||||| || |||||||| |||||||| |||||||| ||||||
Db    296 MPPSESRMSVSSDTDPAAYSVTAVRPHPSISEEWEGFDCLSSPYRAEPPGGGWAGSLGEP 355

Qy    360 PPPPRASLSSDTSALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSDESDSATVYDNCASV 419
          ||| ||||||||| |||||||| |||||||| |||||||||:||||||||| |||||
Db    356 PPPDRASLSSDTSELSYDSVKYTFVVDEHAQLEGVSLRPCFGDHSDESDSATVIDNCASA 415

Qy    420 SSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGL 479
          ||| |||||||| ||:|:|| |:|||||||||:| |||||||| |||||||| |||||
Db    416 SSPKESAIGEEYELAPQPRPPTCMSEDSTPDEPNV-FSKKFLNVQMSGRSRSSSRESFGL 474
```

```
Qy     480 FSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVF 539
           |||: |||| |||| |||||||||| ||||||||| ||||||||| ||||||||| |||||
Db     475 FSCVSNGEEHEQTHTAIFRFVPRHVDELELEVDDWLLVELQAEDAWYEAYNMRTDARGVF 534

Qy     540 PAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRL 599
           |||| |||||||||| ||||||
Db     535 PAYYEIEVTKEPEHFAALAKN---------------------------------------- 555

Qy     600 TVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGF 659
                   ||| ||||||||| ||||: || ||||||||| ||||||||| ||||||||
Db     556 --------SCVGEISVRGVKIHVKADDALEAIGNKCSHFFQKKNISFCGYHLKNNKYFGF 607

Qy     660 ITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYL 710
           | ||||||||| |||||||||:||||||||| ||||||||| |||||||||
Db     608 IMKHPADHRFANHVFVSEDSTQALAESVGRARQQFYKQFVESTCPTEDIYL 658


RESULT 13
US-08-859-201-8
; Sequence 8, Application US/08859201
; Patent No. 5880261
;  GENERAL INFORMATION:
;    APPLICANT:  Waeber, G
;    APPLICANT:  Nicod, P
;    TITLE OF INVENTION:  Transcription Factor Islet-Brain 1 (IB1)
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Luann Cserr
;      STREET:  750 Arimo Avenue
;      CITY:  Oakland
;      STATE:  California
;      COUNTRY:  USA
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/859,201
;      FILING DATE:  20-MAY-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9706731.8
;      FILING DATE:  03-APR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9709920.4
;      FILING DATE:  15-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luann Cserr
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  ME A9702
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  314 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-859-201-8

  Query Match           43.6%;  Score 1655;  DB 2;  Length 314;
  Best Local Similarity  97.1%;  Pred. No. 5.8e-135;
  Matches  305;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy     398 PCFGDYSDESDSATVYDNCASVSSPYESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFS 457
           |||||||||||||||||||| |||||||||||||||:|:|| |||||||||||||||||
Db       1 PCFGDYSDESDSATVYDNCASASSPYESAIGEEYEEAPQPRPPTCLSEDSTPDEPDVHFS 60

Qy     458 KKFLNVFMSGRSRSSSAESFGLFSCIINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLV 517
           |||||||||||||||||||||:||||| |||||||||||||||||||||||||||||||||
Db      61 KKFLNVFMSGRSRSSSAESFGLFSCVINGEEHEQTHRAIFRFVPRHEDELELEVDDPLLV 120

Qy     518 ELQAEDYWYEAYNMRTGARGVFPAYYAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQV 577
           |||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db     121 ELQAEDYWYEAYNMRTGARGVFPAYYAIEVTKEPEHMAALAKNSDWIDQFRVKFLGSVQV 180

Qy     578 PYHKGNDVLCAAMQKIATTRRLTVHFNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSH 637
           |||||||||||||||||||||||||||||||||||||||||||||||||||::||||||||||
Db     181 PYHKGNDVLCAAMQKIATTRRLTVHFNPPSSCVLEISVRGVKIGVKADEAQEAKGNKCSH 240

Qy     638 FFQLKNISFCGYHPKNNKYFGFITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQ 697
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 FFQLKNISFCGYHPKNNKYFGFITKHPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQ 300

Qy     698 FVEYTCPTEDIYLE 711
           ||||||||||||||
Db     301 FVEYTCPTEDIYLE 314


RESULT 14
US-08-859-201-4
; Sequence 4, Application US/08859201
; Patent No. 5880261
;  GENERAL INFORMATION:
;    APPLICANT:  Waeber, G
;    APPLICANT:  Nicod, P
;    TITLE OF INVENTION:  Transcription Factor Islet-Brain 1 (IB1)
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Luann Cserr
;      STREET:  750 Arimo Avenue
;      CITY:  Oakland
;      STATE:  California
;      COUNTRY:  USA
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/859,201
;      FILING DATE:  20-MAY-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9706731.8
;      FILING DATE:  03-APR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9709920.4
;      FILING DATE:  15-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luann Cserr
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  ME A9702
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  270 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-859-201-4

  Query Match           37.6%;  Score 1427;  DB 2;  Length 270;
  Best Local Similarity  99.3%;  Pred. No. 2.3e-115;
  Matches  268;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     442 CLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLFSCIINGEEQEQTHRAIFRFVP 501
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLFSCIINGEEQEQTHRAIFRFVP 60

Qy     502 RHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFPAYYAIEVTKEPEHMAALAKNS 561
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RHEDELXLEVDDPLLVELQAEDYWYEAYNMRTGARGVFPAYYAIEVTKEPEHMAALAKNS 120

Qy     562 DWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRLTVHFNPPSSCVLEISVRGVKIG 621
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     121 DWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRLTVHFNPPSSXVLEISVRGVKIG 180

Qy     622 VKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGFITKHPADHRFACHVFVSEDSTK 681
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 VKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGFITKHPADHRFACHVFVSEDSTK 240
```

```
Qy    682 ALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          |||||||||||||||||||||||||||||||
Db    241 ALAESVGRAFQQFYKQFVEYTCPTEDIYLE 270


RESULT 15
US-09-562-737-31
; Sequence 31, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
;  APPLICANT:  Herz, Joachim
;  APPLICANT:  Gotthardt, Michael
;  TITLE OF INVENTION:  LDL Receptor Signaling Pathways
;  FILE REFERENCE:  UTSW0708
;  CURRENT APPLICATION NUMBER: US/09/562,737
;  CURRENT FILING DATE:  2000-05-01
;  NUMBER OF SEQ ID NOS:  132
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 31
;   LENGTH: 830
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Sequence
US-09-562-737-31

  Query Match             33.7%;  Score 1279.5;  DB 4;  Length 830;
  Best Local Similarity   36.8%;  Pred. No. 6.6e-102;
  Matches  327;  Conservative   94;  Mismatches  193;  Indels  275;  Gaps   32;

Qy     26 HIASPPNFRLTHDISLEEFEDEDLSEITDECGISL-----QC-KDTLSLRPPRAGLLSAG 79
          |  |||  |   |||||||:|||||||||:||: |      | ||:|||
Db     14 HSLSPPGCRPPQDISLEEFDDEDLSEITDDCGLGLSYDSDHCEKDSLSL----------- 62

Qy     80 GGGAGSR------LQAEMLQMDLIDATGDTPGAEDDEEDDDEERAARRPGAGPPKAESGQ 133
            |   :       | :  : ::||   |  |   :||:::||                :
Db     63 --GRSEQPHPICSFQDDFQEFEMID---DHQGGGHEEEEEEEEE--------------E 102

Qy    134 EPASRGQGQSQGQSQGPGS----GDTYRPK---------RPTTLNLFPQVPRSQDTLNNN 180
          |   | ||:: |   ||||    ||:  |          |||||:|      :||:||||
Db    103 EDGDR-QGKAGG---GPGSQALAGDSLIPSPSLEESHKLRPTTLHL--TTLGAQDSLNNN 156

Qy    181 SLG----KKHSWQDRVSRS--SSPLK-------------------------TGEQTPPH 208
          : |       |||: | ||    |||                           | | |
Db    157 NGGFTSAPPSSWQETVLRSPAQEPLKELPAPLLPAEEERHEVQSLARPGCDCEGNQPPE- 215

Qy    209 EHICLSEELPPQS--GPAPTTDRGTSTDSPCR-------RSTATQMAPPGGPPAAPPGGR 259
                 || |  | :|::| |   |           | :: ::: ||       | |
Db    216 ---------PPASSGGASPSSDPGIEADLRSHSSGGHEGRRSSQELSSPGSDSEDAGGAR 266

Qy    260 ------------------GHSHRDRIHYQADVRLEATEEIYLTPVQRPPDAAEP------ 295
                           | |   |  :  :   ||:     :|   |    ||
Db    267 LGRMISSISETELELSSDGGSSSGRSSHLTNSIEEAS-----SPASEPEPEPEPLHEPPR 321

Qy    296 TSAFLP------------PTESRMSVSSDPD------------------PAAYPSTAGRP 325
            ||||             :||   :| | |                   |   |     |
Db    322 RPAFLPVGQDDTNSEYESGSESEPDLSEDADSPWLLSNLVSRMISEGSSPIRCPGQCLSP 381

Qy    326 HPSISEEEEGFDCLSSPERAEPPGGGWRGSL-------GEPPPPPRAS------------ 366
          | : ||       | |:  : |  |     |       | ||| | |
Db    382 APRLPEEAAS-QANSVPQDCQDPEAGPHVELVDMDTLCGPPPPAPAAPRLGPAQPGPCLF 440

Qy    367 LSSDT---------------------------------------------------------- 371
          ||: |
Db    441 LSNPTRDTITPLWATPGRTARPGRSCSAACSEEEEEDEEEDEEDEEDAEDSVVPPGSRTT 500

Qy    372 -------SALSYDSVKYTLVVDEHAQLELVSLRPCFGDYSD-ESDSATVYDNC-ASVSSP 422
                 ::| ||:||||||||||| |||||||| | |  :| | ||     :| ||
Db    501 GSTAPLDASLVYDAVKYTLVVDEHTQLELVSLRRCAGLGNDSEEDS-----SCEASEEEA 555

Qy    423 YESAIGEEYEEAPRPQPPACLSEDSTPDEPDVHFSKKFLNVFMSGRSRSSSAESFGLFSC 482
            : :| :       | |    ||::|| ||: ||||||||||::  ||||| ||||||||

Db    556 GATLLGSD-------QVP----EDASPDSPDLTFSKKFLNVFVNSTSRSSSTESFGLFSC 604

Qy    483 IINGEEQEQTHRAIFRFVPRHEDELELEVDDPLLVELQAEDYWYEAYNMRTGARGVFPAY 542
          : ||||:||||||:|||:||| ||||||:||||:||| : :|:|:  :||||| ||||||:
Db    605 VXNGEERQTHRAVFRFIPRHPDELELDVDDPVLVEAEEDDFWFRGFNMRTGERGVFPAF 664

Qy    543 YAIEVTKEPEHMAALAKNSDWVDQFRVKFLGSVQVPYHKGNDVLCAAMQKIATTRRLTVH 602
          ||  |    : :    ::  |||:| |:|||||:|| |:|| :|||||||||| |:||||
Db    665 YAHAVPGPAKDLLGSKRSPCWVDRFDVQFLGSVEVPCHQGNGILCAAMQKIATARKLTVH 724

Qy    603 FNPPSSCVLEISVRGVKIGVKADDSQEAKGNKCSHFFQLKNISFCGYHPKNNKYFGFITK 662
            ||:|| ||||:||||: :     :     :||||||:||||||| ||:|: |||||||
Db    725 LRPPASCDLEISLRGVKLSLSGGGPE---FQRCSHFFQMKNISFCGCHPRNSCYFGFITK 781

Qy    663 HPADHRFACHVFVSEDSTKALAESVGRAFQQFYKQFVEYTCPTEDIYLE 711
          ||   ||||||||||::| : :| |||||| ::|:: : : |||||||||
Db    782 HPLLSRFACHVFVSQESMRPVARSVGRAFLEYYQEHLAFACPTEDIYLE 830


Search completed: December 24, 2002, 08:51:19
Job time : 22 secs
```